

**UNIVERSIDADE DE LISBOA - FACULDADE DE CIÊNCIAS**

DEPARTAMENTO DE BIOLOGIA ANIMAL



# **Epidemiologia genética do Acidente Vascular Cerebral: identificação de genes envolvidos na susceptibilidade e na recuperação do doente**

Dissertação

Sofia Paulos Pinheiro

**Mestrado em Biologia Humana e Ambiente**

**2012**

**UNIVERSIDADE DE LISBOA - FACULDADE DE CIÊNCIAS**

DEPARTAMENTO DE BIOLOGIA ANIMAL



# **Epidemiologia genética do Acidente Vascular Cerebral: identificação de genes envolvidos na susceptibilidade e na recuperação do doente**

Dissertação

Sofia Paulos Pinheiro

Dissertação orientada pela Doutora Astrid Moura Vicente (Instituto Nacional de Saúde Dr. Ricardo Jorge e Instituto Gulbenkian de Ciência) e pela Prof.<sup>ª</sup> Doutora Maria Teresa Rebelo (Faculdade de Ciências da Universidade de Lisboa)

**Mestrado em Biologia Humana e Ambiente**

**2012**

**Esta dissertação não se encontra escrita segundo o novo acordo ortográfico.**

**As referências bibliográficas seguem a formatação sugerida pela revista científica *Human Genetics*.**

## RESUMO

O Acidente Vascular Cerebral (AVC) é um evento neurológico agudo resultante de uma doença cardiovascular, no qual a afluência de sangue ao cérebro se vê comprometida, seja por um bloqueio ou por uma ruptura num vaso sanguíneo cerebral. Esta doença representa uma das principais causas de morte e morbidade em todo o mundo, pelo que o estudo das componentes ambientais e genéticas nela envolvidas se torna fundamental para a implementação de planos de prevenção que visem minimizar as suas consequências.

Com vista a contribuir para o desenvolvimento do estudo das variantes genéticas implicadas quer na susceptibilidade, quer na recuperação após um AVC, os objectivos deste estudo foram os de, por um lado, estudar a relevância do *Single Nucleotide Polymorphism* (SNP) rs1229984 do gene *ADH1B* sobre o risco de AVC e, por outro, identificar novos polimorfismos envolvidos no processo de recuperação.

Recorrendo a um *TaqMan® Drug Metabolism Genotyping Assay* foi feita uma análise comparativa de dois grupos de indivíduos, um grupo saudável (controlos) e outro de doentes que sofreram AVC (casos) e verificou-se que o SNP rs1229984 do gene *ADH1B*, implicado no metabolismo do álcool, está, de facto, envolvido na susceptibilidade ao AVC. Porém, a mediação desta relação pelo consumo de álcool não se verificou, pelo que se sugere a replicação do estudo.

Tendo por base um rastreio genómico previamente efectuado através de um *Genome Wide Association Study* (GWAS) em *pooled samples*, cujo objectivo era a identificação de genes envolvidos na recuperação após um AVC, realizou-se um novo estudo acerca da recuperação após um AVC, como mesmo objectivo. Seleccionaram-se diversos SNPs segundo critérios bem definidos (e distintos do primeiro estudo), para genotipagem individual, através de ferramentas como *Sequenom iPLEX* e *TaqMan® Pre-Designed SNP Genotyping Assays* e *Restriction Fragment Length Polymorphism* (RFLP), em grupos de indivíduos com recuperação desigual três meses após um AVC. Este estudo permitiu identificar um novo gene candidato envolvido na recuperação, o *CDKAL1* e confirmar o potencial envolvimento do gene *BBS9* no mesmo processo, o qual já tinha sido identificado anteriormente como gene associado à recuperação. Foi também possível aferir alguns aspectos relacionados com a interpretação dos resultados de um GWAS em *pooled samples*, nomeadamente acerca da importância das réplicas da formação dos *pools* e das réplicas de alelotipagem e acerca da necessidade de validação dos resultados por genotipagem individual.

Tendo em conta a identificação de genes candidatos para a recuperação após um AVC, sugere-se que sejam desenvolvidos trabalhos futuros, nomeadamente estudos funcionais, no sentido de compreender as vias através das quais os referidos genes actuarão sobre o processo de recuperação após um AVC. Estudos de replicação em outras populações são também aconselhados, para validação das conclusões tiradas neste estudo.

**Palavras-chave:** AVC; Susceptibilidade; Recuperação; *Single Nucleotide Polymorphism* (SNP); *Genome Wide Association Study* (GWAS); Genes Candidatos.

## ABSTRACT

Stroke is an acute neurological event resulting from a cardiovascular disease in which the blood supply to the brain gets compromised either by a blockage or a rupture in a cerebral blood vessel. This disease is a significant cause of death and disability worldwide, thus the study of genetic and environmental variants involved in stroke susceptibility and outcome becomes essential to allow the implementation of effective prevention plans aiming to minimize its consequences.

To contribute to the study of genetic variants involved in stroke susceptibility or stroke recovery, the main objectives of this study were to determine the relevance of the *ADH1B* gene in stroke risk and to identify new genetic variants involved in stroke recovery.

Using a TaqMan® Drug Metabolism Genotyping Assay we compared genotypes at the rs1229984 within *ADH1B* gene in a group of healthy individuals with a group of stroke patients and found that *ADH1B* gene, which is involved in alcohol metabolism, is associated with stroke susceptibility.

We progressed with the analysis of a previous Genome Wide Association Study (GWAS) for stroke recovery in pooled samples. Several most relevant Single Nucleotide Polymorphisms (SNPs) were selected according to well defined criteria and were individually genotyped in groups of individuals with good and poor outcome at three months evaluation. Genotyping was performed using Sequenom iPLEX, TaqMan® Pre-Designed SNP Genotyping Assays and Restriction Fragment Length Polymorphism (RFLP). This study allowed the identification of a new candidate gene involved in stroke recovery, *CDKAL1*, and also allowed to confirm the involvement of a previously identified candidate gene *BBS9* in the same process. It was also possible to assess some aspects related to the interpretation of the results of a GWAS in pooled samples that are important for the wider use of this technology in genomic studies.

Given the identification of candidate genes, we suggest the development of future studies, including functional studies in order to understand which pathways of recovery the candidate genes affect. Replication in other populations is also suggested as a mean of validating the conclusions drawn in this study.

**Keywords:** Stroke; Susceptibility; Recovery; Single Nucleotide Polymorphism (SNP), Genome Wide Association Study (GWAS); Candidate Genes.

## ATRIBUIÇÕES

O presente trabalho foi desenvolvido no grupo Neurogenética e Saúde Mental (NSM) do Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis (DPS) do Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA). Ao longo de todo o projecto foi possível contar com uma equipa multidisciplinar muito prestável, bem como, pontualmente, com as instalações do Instituto Gulbenkian de Ciência (IGC).

A colheita das amostras de indivíduos com registo de ocorrência de AVC encontrava-se já feita, tendo sido realizada ao abrigo do projecto PECS/T/SAU/179/95, que permitiu concluir acerca da potencialidade da prevenção primária das doenças cardiovasculares (DCVs), nomeadamente no que respeita ao controlo da hipertensão e redução do consumo de álcool e tabaco. Aquando da colheita, todos os indivíduos tinham idade inferior a 65 anos e eram pacientes admitidos em unidades de Neurologia ou Medicina devido à ocorrência de um primeiro AVC (confirmado recorrendo a tomografias axiais computadorizadas e/ou ressonâncias magnéticas). Por AVC entende-se um evento neurológico agudo que resulta num *deficit* neurológico observável por um intervalo de tempo superior a 24h ou que resulta no óbito do indivíduo em menos de 24h. O processo de examinação e selecção dos indivíduos participantes no estudo foi levado a cabo por médicos neurologistas ou de medicina interna de unidades de Neurologia ou Medicina Interna, respectivamente, de 20 hospitais portugueses (continente e ilhas). Simultaneamente à selecção dos pacientes, foram recolhidos dados demográficos, informações relevantes relativas ao estado clínico e ao estilo de vida, características do AVC e sinais e sintomas neurológicos, complicações e intervenções durante a hospitalização e classificação da recuperação com base na modified rankin scale (mRS) (inicialmente na alta após hospitalização e, novamente, aos 3 e 12 meses após o evento). Na maioria dos casos foi colhida ainda uma amostra de sangue periférico.

A colheita das amostras de indivíduos controlo (isto é, sem manifestação de AVC) utilizados encontrava-se igualmente já feita, tendo sido realizada no INSA e no IGC no âmbito de estudos anteriores. Extensa informação acerca da sua identificação e características sociodemográficas, informações relevantes relativas ao estado clínico e ao estilo de vida foi recolhida.

A extracção de DNA a partir do sangue colhido foi feita aquando da colheita e por técnicos especializados, no INSA, através do método de *salting out*, tendo as amostras sido posteriormente quantificadas e armazenadas a 4°C.

A genotipagem dos *single nucleotide polymorphisms* (SNPs) foi feita no decorrer do presente estágio, por mim, através de *TaqMan® Assays* (recorrendo a equipamento do IGC) e de *restriction fragment lenght polymorphisms* (RFLP) (realizados inteiramente na Unidade Laboratorial do DPS, no INSA). Relativamente à genotipagem por Sequenom iPLEX Gold, foi realizada no IGC em conjunto com o Técnico João Costa.

A análise de resultados foi feita maioritariamente por mim, ainda que com muito apoio da Professora Doutora Astrid Moura Vicente, da Doutora Catarina Correia, da Doutora Marta Barreto e do Doutor Baltazar Nunes.

## AGRADECIMENTOS

Ao longo do último ano foram muitas as pessoas que me rodearam e que, de diversas formas, me ajudaram a dar forma a este trabalho. Queria, por isso, deixar aqui o meu agradecimento.

À Doutora Astrid Moura Vicente, minha orientadora, pela oportunidade de integrar o grupo Neurogenética e Saúde Mental, que realiza um trabalho notável acerca de temáticas tão actuais e importantes, como as doenças cardiovasculares e o as doenças do espectro do autismo, e pelas quais tenho um enorme interesse. Obrigada por todas as orientações, críticas e dicas. Graças a este trabalho creio ter crescido muito ao longo do último ano, pessoal e profissionalmente.

À Prof.<sup>a</sup> Doutora Teresa Rebelo, minha co-orientadora e professora, pela atenção e disponibilidade sempre que necessário.

À Isabel Picanço e ao João Costa pela disponibilidade e por toda a ajuda que me deram no laboratório, principalmente aquando do primeiro contacto com novas técnicas de Biologia Molecular.

Ao Doutor Baltazar Nunes e à Doutora Marta Barreto, pela ajuda na análise estatística de alguns resultados.

Às minhas colegas de grupo Inês Conceição, Bárbara Oliveira, Catarina Correia, Ana Rita Tarelho e Joana Coelho que tornaram o meu dia-a-dia mais interessante, quer pelos conhecimentos que me transmitiram, quer por todas as outras conversas que tivemos. Levo comigo boas recordações destes meses!

À Helena Manso e à Vânia Gaio pelas dicas ao longo dos meses.

Às minhas colegas de mestrado, Flávia Leitão, Sara Berguete e Ana Mateus, pelo apoio, pela compreensão, pela motivação e pelo companheirismo no laboratório.

Aos meus amigos de sempre, Joana Marques, Joana Figueiredo, Francisco Martins, Rafael Galupa e Miguel Pinto, que, ainda que por vezes fisicamente longe, estão lá sempre, como as estrelas. Obrigada pela amizade!

Ao Pedro, por todo o amor, apoio e paciência ao longo deste ano. Obrigada pela sua capacidade de me acalmar nos momentos de maior nervosismo (e pelas horas a fio que me ouviu a falar de Biologia Molecular, ainda que sem perceber muito do assunto!).

Aos meus familiares que sempre me acompanham ao longo das etapas importantes da minha vida, com uma palavra de encorajamento e um sorriso de reconhecimento.

Aos meus pais, em particular, que sempre me deram asas para voar e sempre me apoiaram em tudo. O ano da tese não foi excepção e foi, sem dúvida, neles que encontrei os conselhos, a calma e o discernimento para enfrentar os momentos mais difíceis. Obrigada!

## ÍNDICE GERAL

RESUMO .....	iii
ABSTRACT .....	iv
ATRIBUIÇÕES .....	v
AGRADECIMENTOS.....	vi
ABREVIATURAS .....	xii
CAPÍTULO 1. INTRODUÇÃO TEÓRICA GERAL .....	1
1.1. ACIDENTE VASCULAR CEREBRAL (AVC) .....	2
1.1.1. Perspectiva histórica .....	2
1.1.2. Conceito.....	2
1.1.3. Variantes patológicas .....	3
1.1.4. Sintomatologia e Diagnóstico.....	5
1.1.5. Etiologia e Factores de Risco .....	6
1.1.6. Terapêutica.....	9
1.1.7. Epidemiologia .....	9
CAPÍTULO 2. SUSCEPTIBILIDADE AO AVC: A RELEVÂNCIA DO GENE <i>ADH1B</i> .....	11
2.1. INTRODUÇÃO.....	12
2.1.1. Metabolismo do álcool.....	12
2.1.2. Relação entre o consumo de álcool e a doença cardiovascular (DCV) .....	14
2.2. ENQUADRAMENTO E OBJECTIVOS.....	15
2.2.1. Enquadramento.....	15
2.2.2. Objectivos.....	16
2.3. METODOLOGIA.....	16
2.3.1. Caracterização da amostra .....	16
2.3.2. Processamento da amostra .....	16
2.3.3. Genotipagem individual .....	17
2.3.4. Análise estatística dos resultados .....	19
2.4. RESULTADOS.....	19
2.5. DISCUSSÃO .....	24



CAPÍTULO 3. IDENTIFICAÇÃO DE VARIANTES GENÉTICAS QUE INFLUENCIAM A RECUPERAÇÃO APÓS UM AVC ISQUÉMICO.....	28
3.1. INTRODUÇÃO.....	29
3.1.1. Escalas de classificação da recuperação.....	30
3.1.2. Factores envolvidos na recuperação.....	30
3.1.3. Prevenção secundária .....	32
3.2. ENQUADRAMENTO E OBJECTIVOS.....	32
3.2.1. Enquadramento.....	32
3.2.2. Objectivos.....	34
3.3. METODOLOGIA.....	35
3.3.1. Caracterização da amostra .....	35
3.3.2. Processamento da amostra.....	35
3.3.3. Selecção dos SNPs a analisar .....	36
3.3.4. Genotipagem individual .....	38
3.3.5. Análise estatística dos resultados .....	46
3.4. RESULTADOS.....	47
3.4.1. Validação dos resultados do GWAS em <i>pooled samples</i> por genotipagem individual .....	48
3.4.2. Análise de associação dos 24 SNPs validados na amostra combinada .....	50
3.5. DISCUSSÃO .....	51
CAPÍTULO 4. CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS .....	57
REFERÊNCIAS BIBLIOGRÁFICAS .....	59
ANEXOS.....	64
FIGURAS SUPLEMENTARES.....	65
TABELAS SUPLEMENTARES.....	68

## ÍNDICE DE TABELAS

Tabela 1 – Tabela-resumo dos sintomas e respectivos prognósticos associados a um AVC .....	6
Tabela 2 - Relação observada entre um determinado genótipo e o nível de metabolização de etanol .....	13
Tabela 3 – Caracterização da amostra populacional utilizada no estudo da relação entre o gene <i>ADH1B</i> e a susceptibilidade ao AVC .....	17
Tabela 4 – Condições da mistura de reacção utilizada na execução do ensaio <i>TaqMan® Drug Metabolism Genotyping Assay</i> para a genotipagem do SNP rs1229984 .....	18
Tabela 5 – Programa utilizado na execução do ensaio <i>TaqMan® Drug Metabolism Genotyping Assay</i> para a genotipagem do SNP rs1229984 .....	18
Tabela 6 – Resultados da genotipagem individual do SNP rs1229984.....	20
Tabela 7 – Análise de associação entre o genótipo e o AVC .....	20
Tabela 8 – Análise de associação entre o genótipo e o consumo de álcool .....	21
Tabela 9 – Análise de associação entre o consumo de álcool e o AVC .....	21
Tabela 10 - Análise de associação entre o genótipo e o consumo de álcool, considerando os grupos de casos e controlos separadamente .....	22
Tabela 11 – Efeito modificador do consumo de álcool sobre a associação entre o genótipo e o AVC .....	23
Tabela 12 - Efeito modificador do SNP sobre a associação entre o consumo de álcool e o AVC .....	24
Tabela 13 – <i>Modified Rankin Scale</i> (mRS) .....	31
Tabela 14 – Caracterização da amostra combinada utilizada no estudo de variantes genéticas associadas à recuperação após um AVC .....	35
Tabela 15 – Caracterização dos SNPs genotipados individualmente .....	37
Tabela 16 – Resumo dos SNPs genotipados recorrendo a cada uma das três técnicas utilizadas para o efeito .....	39
Tabela 17 – Quantidades e concentrações dos reagentes no volume final de reacção de PCR para genotipagem com RFLP .....	41
Tabela 18 – Programas utilizados nas PCRs dos SNPs a analisar por RFLP.....	41
Tabela 19 - Características e condições das enzimas utilizadas na genotipagem dos 4 SNPs por RFLP .....	42
Tabela 20 – Quantidades e concentrações dos reagentes no volume final de reacção de PCR de Sequenom iPLEX Gold.....	43
Tabela 21 – Programa utilizado nas PCR dos SNPs a analisar por Sequenom iPLEX Gold .....	43
Tabela 22 – Quantidades dos reagentes na mistura de reacção de neutralização de Sequenom iPLEX Gold .	44
Tabela 23 – Condições da incubação da fase de neutralização de Sequenom iPLEX Gold .....	44
Tabela 24 – Quantidades dos reagentes na mistura de reacção de extensão de Sequenom iPLEX Gold.....	44
Tabela 25 – Programa utilizado na etapa de extensão de Sequenom iPLEX Gold .....	45
Tabela 26 – Condições da mistura de reacção utilizada na execução dos ensaios <i>TaqMan® SNP Pre-Designed Genotyping Assay</i> .....	45

Tabela 27 – Programa utilizado na execução dos ensaios <i>TaqMan® SNP Pre-Designed Genotyping Assay</i> para a genotipagem do SNP rs1229984 .....	46
Tabela 28 – SNPs genotipados recorrendo à técnica <i>TaqMan® SNP Pre-Designed Genotyping Assay</i> com respectivos polimorfismos e fluoróforos acoplados a cada alelo .....	46
Tabela 29 – Resultados da validação da alelotipagem para cada SNP por genotipagem individual .....	48
Tabela 30 – Resultados do teste de associação dos SNPs validados na amostra combinada (N=414).....	52

## TABELAS SUPLEMENTARES

Tabela suplementar 1 – Sequências dos <i>primers forward</i> e <i>reverse</i> utilizados nas PCRs dos SNPs a analisar por RFLP .....	68
Tabela suplementar 2 – Sequências dos <i>primers forward</i> e <i>reverse</i> utilizados na PCR dos SNPs a analisar por Sequenom iPLEX .....	68
Tabela suplementar 3 – Sequências das sondas de extensão utilizadas em Sequenom iPLEX Gold .....	70

## ÍNDICE DE FIGURAS

Figura 1 – Cascata de processos biológicos que ocorrem após um evento isquêmico.....	4
Figura 2 – Dimensão e posição relativas dos sete genes <i>ADH</i> humanos, ao longo do braço longo do cromossoma 4 (4q).....	12
Figura 3 – Representação esquemática da proteína ADH1B codificada por cada um dos três alelos do gene <i>ADH1B</i> .....	13
Figura 4 – Gráfico representativo da relação entre o consumo de álcool e a mortalidade.....	14
Figura 5 – Representação esquemática do PCR realizado no <i>TaqMan Drug Metabolism Genotyping Assay</i> aquando da genotipagem do SNP rs1229984 .....	18
Figura 6 – Representação esquemática das três fases da técnica de <i>TaqMan Drug Metabolism Genotyping Assay</i> .....	19
Figura 7 – Limiar da corrente sanguínea cerebral para a disfunção celular e morte.....	29
Figura 8 – <i>Workflow</i> do trabalho efectuado previamente .....	33
Figura 9 – Gráfico representativo das maiores diferenças entre as frequências alélicas nos dois grupos (mRS=0 e mRS≥3) .....	36
Figura 10 – Respresentação esquemática das fases da técnica Sequenom iPLEX Gold.....	42
Figura 11 – Diferenças entre as frequências alélicas dos grupos calculadas após alelotipagem e genotipagem individual. ....	50

## FIGURAS SUPLEMENTARES

Figura Suplementar 1 - Gráfico da distribuição do <i>top 1000</i> (CV<10%). SNPs com maiores diferenças entre as frequências alélicas dos grupos (mRS=0 e mRS≥3) calculadas após a alelotipagem dos 262 264 analisados no GWAS.....	65
Figura Suplementar 2 – Gráfico da distribuição dos 1000 SNPs com <i>p-values</i> mais baixos, após a realização de um teste <i>t</i> de Student para verificar a significância das diferenças entre as frequências alélicas nos dois grupos (boa recuperação – mRS=0 e má recuperação – mRS≥3) .....	66
Figura Suplementar 3 – Representação esquemática dos locais de corte das enzimas de restrição utilizadas nas sequências amplificadas por PCR e padrão de fragmentos associados a cada genótipo.....	67

## ABREVIATURAS

3' – Região não traduzida 3'UTR (*Three prime Untranslated Region*)

5' – Região não traduzida 5'UTR (*Five prime Untranslated region*)

A – adenina

ADH – *Alcohol dehydrogenases*

ADH – gene codificante de *Alcohol dehydrogenases*

ADH1B – *Alcohol dehydrogenases 1B*

ADH1B – gene codificante de *Alcohol dehydrogenases 1B*

ADH1C – *Alcohol dehydrogenases 1C*

ADH1C – gene codificante de *Alcohol dehydrogenases 1C*

ALDH – *Aldehyde dehydrogenases*

ALDH – gene codificante de *Aldehyde dehydrogenases*

ALDH2 – *Aldehyde dehydrogenase 2 family*

ALDH2 – gene codificante de *Aldehyde dehydrogenase 2 family*

ALOX5AP – gene codificante de *Arachidonate 5-lipoxygenase-activating protein*

APOE – gene codificante de *Apolipoprotein E*

Arg – Arginina

AVC – Acidente vascular cerebral

BBS9 – gene codificante de *Bardet-Biedl syndrome 9*

BI – Barthel index

Bp – par de bases (*base pair*)

BSA – *Bovine Serum Albumin*

C – citosina

C4orf19 – *chromosome 4 open reading frame 19*

CADASIL – *cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy*

cAMP – adenosina 3'-5'-monofosfato cíclico

CDK5RAP1 – gene codificante de *CDK5 regulatory subunit associated protein 1*

CDKAL1 – gene codificante de *CDK5 regulatory subunit associated protein 1-like 1*

CENPP – gene codificante de *Centromere protein P*

CEU – população caucasiana europeia e de descendência europeia no Utah (EUA)

CNTP5 – gene codificante de *Contactin-associated protein-like 5 precursor*

COX-2 – gene codificante de *Cyclooxygenase-2*

Cr – cromossoma

CRP – gene codificante de *C-reactive protein*

CSMD1 – gene codificante de *CUB and Sushi multiple domains 1*

CV – coeficiente de variação

CVT – Trombose venosa cerebral (*cerebral venous thrombosis*)

*DCC* – gene codificante de *Deleted in colorectal carcinoma*  
*DCV* – doença cardiovascular  
*DNA* – ácido desoxirribonucleico (*deoxyribonucleic acid*)  
*dNDP* – desoxirribonucleótidos difosfatados  
*dNTP* – desoxirribonucleótidos trifosfatados  
*DP* – desvio padrão  
*DPS* – Departamento de Promoção da Saúde e Doenças Crônicas  
*FCT* – Fundação para a Ciência e Tecnologia  
*G* – guanina  
*GLIS3* – gene codificante de *GLIS family zinc finger 3*  
*Glu* – glutamato  
*GP1IIa* – gene codificante de *Glycoprotein 1IIa*  
*GRM7* – gene codificante de *Glutamate receptor, metabotropic 7*  
*GSTO2* – gene codificante de *Glutathione S-transferase omega 2*  
*GWAS* – *Genome wide association study*  
*HDL* – lipoproteína de alta densidade (*high-density lipoprotein*)  
*His* – histidina  
*HSPA4* – gene codificante de *Heat shock 70kDa protein 4*  
*HWE* – Equilíbrio de Hardy Weinberg (*Hardy Weinberg equilibrium*)  
*ICH* – Hemorragia intracerebral (*intracerebral hemorrhage*)  
*IL1B* – gene codificante de *Interleukin 1, beta*  
*IL6* – gene codificante de *Interleukin 6*  
*IMC* – índice de massa corporal  
*INSA* – Instituto Nacional de Saúde Dr. Ricardo Jorge  
*Kb* – Quilo bases (*Kilo base pair*)  
*KCNB1* – gene codificante de *Potassium voltage-gated channel, Shab-related subfamily, member 1*  
*LD* – desequilíbrio de *linkage* (*linkage disequilibrium*)  
*Lys* – lisina  
*MAF* – *minor allele frequency*  
*MALDI-TOF* – *matrix assisted laser desorption/ionization – time-of-flight*  
*Mb* – Mega bases (*Mega base pair*)  
*MgCl<sub>2</sub>* – cloreto de magnésio  
*MMP* – *Matrix metalloproteinase*  
*MMP-2* – gene codificante de *Matrix metalloproteinase 2*  
*MMP3* – gene codificante de *Matrix metalloproteinase 3*  
*MMP-9* – gene codificante de *Matrix metalloproteinase 9*  
*MNP* – *multi-nucleotide polymorphism*  
*MPO* – gene codificante de *Myeloperoxidase*

mRS – *modified Rankin scale*

MTHFR – gene codificante de *Methylenetetrahydrofolate reductase*

MYST4 – gene codificante de *MYST histone acetyltransferase monocytic leukemia 4*

NEB – *New England Biolabs*

NIHSS – *National Institutes of Health stroke scale*

NOS – *Nitric oxide synthase*

NOS1 – gene codificante de *Nitric oxide synthase 1*

NSM – Neurogenética e Saúde Mental

OR – *odds ratio*

P – *p-value*

PCR – reacção de polimerização em cadeia (*polymerase chain reaction*)

PDE4D – gene codificante de *Phosphodiesterase 4D*

PLCB1 – gene codificante de *Phospholipase C, beta 1*

Pooled samples – amostras agrupadas

Primers – *oligonucleótidos de iniciação*

PTGS2 – gene codificante de *Prostaglandin-endoperoxide synthase 2*

RFLP – *restriction fragment lenght polymorphism*

RM – ressonância magnética

rx – reacção

SAH – Hemorragia subaracnóide (*subarachnoid hemorrhage*)

SAP – fosfatase alcalina de camarão (*shrimp alkaline phosphatase*)

SEN5 – gene codificante de *SUMO1/sentrin specific peptidase 5*

SERPINE1 – gene codificante de *Plasminogen activator inhibitor type 1*

SIS – *stroke impact scale*

SNP – *Single Nucleotide Polymorphism*

T – timina

TAC – tomografia axial computadorizada

TIA – acidente isquémico transitório (*transient ischemic attack*)

Tm – temperatura de *melting* (*melting temperature*)

TMEM178 – gene codificante de *Transmembrane protein 178A*

TNF – gene codificante de *Tumor necrosis factor*

tPA – plasminogénico activador tecidual (*tissue plasminogen activator*)

U – unidades (*unit*)

VPS53 – gene codificante de *Vacuolar protein sorting 53 homolog (S. cerevisiae)*

# **CAPÍTULO 1. INTRODUÇÃO TEÓRICA GERAL**



## 1.1. ACIDENTE VASCULAR CEREBRAL (AVC)

### 1.1.1. Perspectiva histórica

Foi há cerca de 2400 anos, no século IV a.C., que o actualmente denominado Acidente Vascular Cerebral (AVC) foi descrito pela primeira vez. A doença foi identificada por Hipócrates, após a análise de diversos casos de pacientes que tinham um quadro clínico de paralisias súbitas. Inicialmente, esta patologia foi denominada apoplexia; no entanto, à medida que se foram descobrindo diversas causas por detrás das paralisias, este termo tornou-se inespecífico, uma vez que não as discriminava, indicando unicamente que alguém teria sofrido uma qualquer paralisia (NINDS, 2004; Warlow *et al.*, 2007).

Durante muitos anos, a Medicina nada soube fazer para ajudar a resolver as consequências dos AVCs, mas actualmente têm vindo a ser desenvolvidas terapias eficientes que podem ajudar os pacientes a recuperar totalmente. Os progressos nesta área iniciaram-se graças a Johann Jacob Wepfer que, no século XVII, identificou em autópsias sinais de hemorragias cerebrais em pacientes cuja morte se devera a apoplexia. Subsequentemente a esta descoberta foram identificadas pela primeira vez as maiores artérias que irrigam o cérebro e, assim, justificadas as hemorragias já observadas. Por outro lado, este médico propôs ainda que a apoplexia, para além de ser provocada então pela ruptura das artérias cerebrais, poderia também ter na sua origem um bloqueio nas mesmas. À luz dos novos conhecimentos, a apoplexia causada pela ruptura ou bloqueio de artérias cerebrais passou a denominar-se doença cerebrovascular (NINDS, 2004; Warlow *et al.*, 2007).

Após o contributo de Johann Jacob Wepfer, foram muitos os profissionais de saúde que se dedicaram ao estudo dos AVCs. Nas duas últimas décadas têm sido alcançados objectivos importantes, já que se conseguiram identificar factores de risco, desenvolver terapias farmacológicas e delinear planos de prevenção efectivos (NINDS, 2004).

Estudos em modelos animais demonstram a importância de uma actuação rápida após a ocorrência de um AVC, pois os danos cerebrais ocorrem em poucos minutos e tornam-se permanentes se não tratados em poucas horas (NINDS, 2004).

### 1.1.2. Conceito

O AVC é descrito actualmente como um evento neurológico agudo que resulta de uma doença cardiovascular e que conduz ao aparecimento de sequelas observáveis por mais de 24h ou, em casos mais graves, origina mesmo a morte. O AVC é caracterizado por um comprometimento da afluência de sangue ao cérebro e mediante ocorra uma oclusão ou uma ruptura num vaso sanguíneo cerebral, assim se pode classificar o evento em isquémico ou hemorrágico, respectivamente (Brass, 1992; Domingues-Montanari *et al.*, 2008; Silverman *et al.*, 2011; Warlow *et al.*, 2007).

Existem múltiplos processos complexos e vias de interacção na origem de um AVC (Lanktree *et al.*, 2010). Porém, ainda que a etiologia dos AVCs não seja conhecida na íntegra, é sabido que a severidade das suas consequências dependerá de diversos factores, entre os quais a região cerebral afectada, o fenómeno na sua origem (bloqueio ou ruptura do vaso sanguíneo) e o intervalo de tempo decorrido desde o evento até à

obtenção de assistência médica (quanto maior o intervalo de tempo que uma determinada região cerebral estiver sem fluxo sanguíneo, mais grave o quadro clínico do paciente, bem como o carácter permanente das lesões) (Brass, 1992).

### 1.1.3. Variantes patológicas

Tal como já foi referido, os AVCs podem ser muito heterogéneos, pelo que é comum classificarem-se em diversas variantes patológicas. Segundo Warlow e colaboradores (2003) seria possível discriminar três variantes patológicas: AVCs isquémicos, hemorragias intracerebrais (ICHs) e hemorragias subaracnóides (SAHs). No entanto, há autores que optam por uma classificação mais simples, subdividindo os AVCs geralmente em isquémicos ou hemorrágicos (incluindo nesta última categoria as ICHs e as SAHs, enumeradas anteriormente) (Saenger & Christenson, 2010). Existem ainda, porém, autores que, para além dos tipos de AVCs já mencionados, defendem a discriminação de trombozes venosas cerebrais (CVTs) e AVCs da medula espinal (Amarenco *et al.*, 2009).

No caso deste trabalho, considerar-se-á apenas a classificação dos AVCs em isquémicos e hemorrágicos, dado não haver informações mais específicas acerca dos pacientes que constituem a amostra que será analisada.

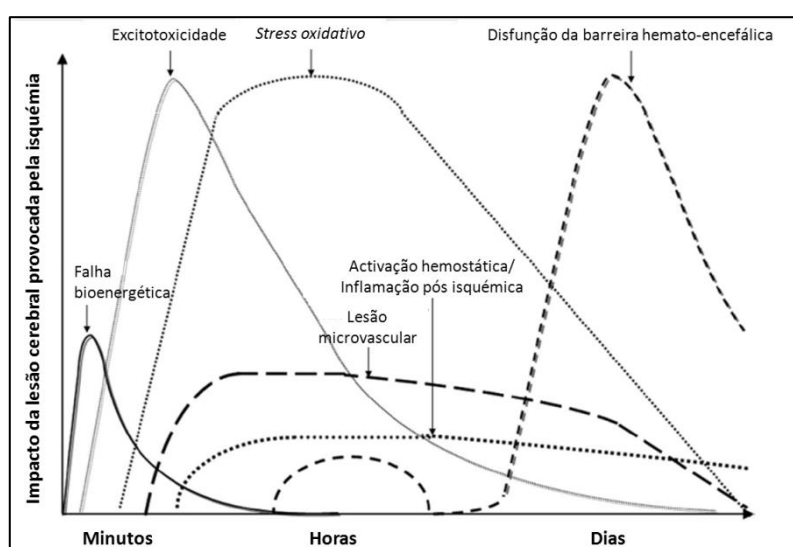
#### AVCs isquémicos

Os AVCs isquémicos constituem cerca de 80% do total de AVCs diagnosticados, sendo a variante patológica mais comum (Guo *et al.*, 2010).

A isquémia é o resultado da diminuição ou cessação da irrigação sanguínea de um tecido, o que resulta na privação de oxigénio e nutrientes nas quantidades adequadas. Como resultado, desenvolvem-se processos de lesão que poderão ser ou não reversíveis. Nestes casos, a rapidez na obtenção de cuidados médicos torna-se fundamental, a fim de minimizar a probabilidade de permanência das lesões (NINDS, 2004). Ainda que por vezes ocorram eventos isquémicos de mais curta duração (TIAs – *transient ischemic attacks*) e que, por isso, não chegam a ser considerados AVCs, o seu diagnóstico é igualmente importante, já que representam um factor de risco para o desenvolvimento de futuros AVCs (Brass, 1992).

A maioria dos AVCs isquémicos ocorre devido à formação de trombos que bloqueiam alguns vasos sanguíneos cerebrais. A capacidade de coagulação do sangue, ao mesmo tempo que é fundamental para a manutenção da saúde, nomeadamente no que respeita aos processos de cicatrização, pode também ser prejudicial, quando os trombos se desenvolvem desnecessariamente ou se deslocam para vasos sanguíneos de baixo calibre (Brass, 1992; NINDS, 2004). Assim, os AVCs isquémicos podem subdividir-se em trombozes intracranianas ou embolismos extracranianos, os mais comuns, consoante o trombo tenha origem num vaso sanguíneo cerebral ou tenha migrado desde outro local do organismo, respectivamente (Saenger & Christenson, 2010; Silverman *et al.*, 2011).

Em termos de fisiopatologia, há diferenças evidentes entre AVCs isquémicos e hemorrágicos. Relativamente aos AVCs isquémicos, sabe-se que diminuído ou interrompido o fluxo de sangue num determinado local é desencadeada uma cascata de processos biológicos (Figura 1) conducentes a lesões. Entre os referidos processos pode enumerar-se a hipoperfusão de nutrientes e oxigénio; falhas bioenergéticas (uma vez que as bombas iónicas das membranas celulares perdem igualmente a sua funcionalidade); excitotoxicidade (devido à libertação excessiva de neurotransmissores excitatórios como o glutamato); stress oxidativo (devido à produção de espécies reactivas de oxigénio); disfunção da barreira hemato-encefálica, danos microvasculares, inflamação e necrose de células neuronais, gliais e endoteliais (devido à libertação, entre outros, de metaloproteases da matriz – MMPs); activação hemostática, entre outros (Deb *et al.*, 2010; Saenger & Christenson, 2010).



**Figura 1** – Cascata de processos biológicos que ocorrem após um evento isquémico (Saenger & Christenson, 2010).

### AVC hemorrágicos

Contrastando com os AVCs isquémicos, os AVCs hemorrágicos constituem apenas cerca de 20% dos AVCs totais. Podendo sub-classificar-se em ICHs (75%) e SAHs (25%), como referido anteriormente, a designação depende essencialmente do tipo de vaso sanguíneo em que ocorre a ruptura que origina a hemorragia, da sua localização e dos factores na sua origem. Geralmente, as ICHs têm maior incidência nas artérias de menor calibre localizadas no interior do cérebro e devem-se essencialmente a quadros hipertensivos. As SAHs, por sua vez, são o tipo de AVC menos comum e devem-se à ruptura de aneurismas ou a malformações vasculares que originam hemorragias na região subaracnóide (Brass, 1992; Deb *et al.*, 2010; Warlow *et al.*, 2003).

Dependendo do volume de sangue vertido para a região cerebral, assim a gravidade do AVC será distinta. Os danos causados resultarão de uma cascata de processos que ocorrem durante e após a hemorragia. Entre outras vias alteradas, observa-se uma irrigação sanguínea deficiente, o que resulta na indução de um estado de hipoxia; por outro lado, o contacto do sangue com o parênquima cerebral e com a vasculatura desencadeará processos de inflamação nos tecidos; no caso de uma hemorragia prolongada, também a

pressão intracraniana tenderá a aumentar, comprometendo de forma agravada a irrigação sanguínea a outras regiões do cérebro (Brass, 1992; Deb *et al.*, 2010; Xi *et al.*, 2006).

O AVC hemorrágico é, portanto, mais grave que um AVC isquémico e apresenta uma maior taxa de mortalidade associada. É sabido que um volume hemorrágico superior a 150mL tende a ser fatal, enquanto os derrames de volume igual ou inferior a 140mL têm um prognóstico variável. De facto, é impossível prever danos cerebrais secundários causados pelo hematoma gerado, quer por se desconhecer o seu volume, quer por ser difícil prever quais as regiões cerebrais que se verão afectadas (Deb *et al.*, 2010; Xi *et al.*, 2006).

#### 1.1.4. Sintomatologia e Diagnóstico

Um AVC desencadeia um conjunto de sintomas típicos e limitativos que conduzem o paciente à necessidade de recorrer a cuidados médicos. Mediante a região cerebral afectada, assim os sintomas e a sua severidade poderão variar (

Tabela 1). O sintoma mais associado à distinção entre um AVC hemorrágico e isquémico é a ocorrência, no primeiro, de uma dor de cabeça forte e repentina. Todos os outros sintomas (fraqueza motora, ataxia, perda sensorial, alteração da acuidade visual e da percepção visuoespacial, alterações ao nível da linguagem e da produção de discurso, défices cognitivos e défices comportamentais) poderão ocorrer em ambos os tipos de AVCs (Silverman *et al.*, 2011).

Não existe, portanto, uma linearidade nos sintomas de AVCs isquémicos e hemorrágicos, estando um correcto diagnóstico dependente de uma análise profissional cuidada e específica de cada paciente.

Apesar dos sintomas característicos de um AVC, é fundamental confirmar a sua origem vascular, já que por vezes existem patologias como tumores, enxaquecas e traumatismos que podem mimetizar alguns dos referidos sintomas. Desta forma, o diagnóstico feito a partir da conjugação da história clínica do paciente, de um exame neurológico e de um estudo imagiológico é fundamental e torna-se muito útil na orientação do tratamento na fase aguda, assim como no estabelecimento de um prognóstico e de um plano de prevenção secundária adequado (Davenport & Dennis, 2000; Silverman *et al.*, 2011).

Os exames imagiológicos mais frequentemente utilizados são a tomografia axial computadorizada (TAC) e a ressonância magnética (RM). A TAC é o método mais largamente utilizado e permite essencialmente, numa fase precoce, excluir a possibilidade dos sintomas apresentados serem originados por patologias cujos sintomas mimetizem AVCs. A partir de algumas horas após o evento, permite observar a isquémia ou hemorragia, sempre que estas existam. Para a observação de uma hemorragia, a TAC apenas é efectiva até uma semana após o evento, uma vez que decorrido esse intervalo de tempo as lesões que se observam assemelham-se às deixadas por um AVC isquémico. A RM é, por outro lado, uma técnica mais sofisticada. Contudo, em contexto de diagnóstico a sua eficiência é semelhante à da TAC. Uma hemorragia, nas primeiras horas, não é indubitavelmente detectável e um AVC isquémico não é facilmente diagnosticado por esta técnica até alguns dias após ter ocorrido. Porém, a RM, ao contrário da TAC, permite observar lesões anteriores provocadas por hemorragias antigas, o que pode ser importante para compreender e interpretar o historial do doente (Davenport & Dennis, 2000; Ringelb *et al.*, 2011; Warlow *et al.*, 2003).

**Tabela 1** – Tabela-resumo dos sintomas e respectivos prognósticos associados a um AVC (Silverman *et al.*, 2011)

Sintoma	Região cerebral afectada	Consequências	Prognóstico
<b>Fraqueza motora</b>	Córtex periférico	Paresia (fraqueza localizada) afectando a face e/ou braço e/ou mão e/ou dedos	A recuperação mais acentuada ocorre nos primeiros 1-3 meses após o AVC. A reabilitação até aos 9 meses revela-se benéfica.
	Tronco cerebral, regiões sub-corticais e/ou proximais	Hemiparesia (fraqueza generalizada) afectando a face e/ou braço e/ou perna de um mesmo lado do corpo	
<b>Ataxia</b>	Hemisfério cerebelar	Descoordenação motora	A recuperação a longo-prazo tem excelente prognóstico.
<b>Perda sensorial</b>	Tálamo lateral, tronco cerebral, região lenticulocapsular	Alteração da percepção de estímulos dolorosos e térmicos, agrafestesia (incapacidade de perceber a escrita sobre a pele, de olhos fechados) e aestereognosia (incapacidade de reconhecer ou identificar a forma e os contornos dos objetos através do tacto)	
<b>Alteração da acuidade visual</b>	Embolismo da artéria carótida	Amaurose (perda parcial ou total da visão)	A recuperação é espontânea e ocorre até 3 meses após o AVC
	Vias visuais nos hemisférios cerebrais	Perda visual de um hemiscampo	
	Lobos occipitais (ambos)	Cegueira cortical	
	Regiões posteriores do cérebro	Diplopia (visão dupla)	
	Região frontal de um hemisfério	Paralisia ocular	
<b>Alteração da percepção visuoespacial</b>	Hemisfério cerebral direito	Alteração da percepção da região esquerda do corpo Anosognosia (incapacidade de ter consciência da própria doença)	
<b>Alterações ao nível da linguagem e da produção de discurso</b>	Tronco cerebral e cerebelo e regiões que provocam fraqueza localizada na face	Disartria (incapacidade de articular as palavras de forma correcta)	A recuperação pode levar de meses a anos
	Regiões do hemisfério esquerdo	Afasia (deterioração da função da linguagem)	
<b>Défices cognitivos</b>		Afasia e inatenção	
<b>Défices comportamentais</b>		Apraxia (incapacidade de executar movimentos e gestos precisos), perda de memória, demência, fadiga, depressão e outras patologias psiquiátricas	

### 1.1.5. Etiologia e Factores de Risco

Os factores etiológicos de um AVC isquémico são geralmente distintos dos de um AVC hemorrágico. Todavia, ambas as variantes patológicas da doença são, na maioria dos casos, multifactoriais, isto é, desencadeadas pela combinação de factores genéticos e não genéticos (factores de risco). Podendo a interacção destes factores ser muito variada, por vezes não é possível identificar a causa exacta que desencadeou um evento (Davenport & Dennis, 2000).

Relativamente aos AVCs isquémicos, a causa mais comum de trombozes intracranianas é a aterosclerose, doença inflamatória crónica em que se formam placas ateroscleróticas nas paredes dos vasos sanguíneos, as quais, pelas propriedades dos seus constituintes, potenciam o desenvolvimento de trombos à sua superfície.

Já os embolismos extracranianos têm como principal causa patologias cardíacas ou da artéria carótida que potenciam igualmente o desenvolvimento de trombos que irão migrar eventualmente para o cérebro. Contudo, existem outras causas para os AVCs isquémicos: por vezes, ao invés de um trombo, num vaso aterosclerótico forma-se uma placa de ateroma de tal forma proeminente que causa a estenose do vaso, impedindo o fluxo do sangue (Brass, 1992; NINDS, 2004; Saenger & Christenson, 2010). Raras vezes, e contrariando a sua natureza multifactorial, os AVCs isquémicos podem resultar de um quadro patológico originado por doenças monogénicas. É o caso dos que advêm de doenças congénitas arteriais de pequenos/médios/grandes vasos (como por exemplo a displasia fibromuscular e a CADASIL – *cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy*), de coagulopatias, de doenças metabólicas relacionadas com arteriopaties e de problemas auto-imunes como as vasculites (Amarenco *et al.*, 2009; Deb *et al.*, 2010).

No que respeita aos AVCs hemorrágicos, as ICHs são essencialmente desencadeadas pela hipertensão. A elevada pressão sanguínea exercida nos vasos de menor calibre acaba por desgastá-los e originar a sua ruptura. Contudo, este factor não é exclusivo na etiologia das ICHs: neoplasias, eventos traumáticos, distúrbios hemorrágicos, angiopatia amilóide, malformações vasculares ou o consumo de substâncias ilícitas opiáceas são também factores potencialmente desencadeantes de uma ICH. Como factores agravantes podem ainda enumerar-se a diabetes e a aterosclerose. As SAHs, por sua vez, são o tipo de AVC menos comum e devem-se geralmente à ruptura de aneurismas ou a malformações vasculares, como referido anteriormente (Brass, 1992; Deb *et al.*, 2010; Warlow *et al.*, 2003). À semelhança do que ocorre nos AVCs isquémicos, também os hemorrágicos podem ser, embora muito raramente, desencadeados por doenças monogénicas. Estas são geralmente formas de amiloidose autossómicas dominantes que afectam a vasculatura cerebral, provocando AVCs recorrentes e a morte prematura do paciente (Guo *et al.*, 2010).

À excepção das patologias monogénicas que desencadeiam AVCs, todas as outras possíveis causas na origem destes eventos resultam da interacção entre factores genéticos e não genéticos.

Tendo em conta uma abordagem no sentido de identificar variantes genéticas envolvidas na susceptibilidade ao AVC, diversos estudos têm sido desenvolvidos. Alguns encontram associação de genes envolvidos em processos inflamatórios, como o *CRP*, o *MPO* e o *IL6*, com um risco aumentado de AVC, o que, de facto, é espectável, uma vez que processos inflamatórios ao nível do sistema nervoso central estão frequentemente relacionados com patologias cerebrais como os AVCs (Guo *et al.*, 2010; Manso *et al.*, 2011). Outros relacionam o papel das enzimas sintases do monóxido de azoto (NOS) com as doenças cardiovasculares (como a aterosclerose e, consequentemente, o AVC), nomeadamente pelo seu potencial envolvimento na vasodilatação e manutenção da tensão arterial, e mostram uma associação entre o gene *NOS1* e a susceptibilidade ao AVC (Manso *et al.*, 2012). Outros ainda mostram uma associação entre genes envolvidos na via de síntese dos leucotrienos e o risco de AVC. É o caso do gene *ALOX5AP* que codifica proteína activadora da 5-lipoxigenase e que, simultaneamente foi identificado como implicado no risco de desenvolvimento de aterosclerose (Gulcher *et al.*, 2005). A análise de factores envolvidos na via de síntese dos leucotrienos é pertinente no âmbito da análise etiológica de AVCs, pois estes lípidos estão envolvidos

em processos inflamatórios crónicos (referidas anteriormente como relacionados com o risco de AVC) e no aumento da permeabilidade vascular, o que favorece o edema na zona afectada pelo AVC (Guo *et al.*, 2010; Matarin *et al.*, 2010). O gene que codifica a fosfodiesterase 4D (*PDE4D*), proteína responsável pela degradação do segundo mensageiro cAMP que, por sua vez, é responsável pela transdução de sinal em células vasculares, entre outras, tem-se também mostrado associado ao risco de AVC isquémico (Guo *et al.*, 2010). Segundo Matarin e colaboradores (2010), os genes codificantes das proteínas metilenotetrahidrofolato redutase (MTHFR), protrombina G20210A e enzimas conversoras da angiotensina estão também associados ao risco de AVC.

Genes envolvidos em outras vias potencialmente relacionadas com o AVC têm também sido objecto de estudo. É o caso do *SERPINE1*, cuja proteína codificada está envolvida no sistema hemostático, do *APOE*, determinante no metabolismo de lipoproteínas, do *TNF*, associado a processos inflamatórios e do *MMP-3*, relacionado com a manutenção das matrizes extracelulares. Contudo, os resultados têm-se mostrado inconclusivos, no sentido em que nem sempre tem sido possível replicar com sucesso estudos que remetem para uma associação entre os genótipos e os fenótipos de AVC (Guo *et al.*, 2010). O esclarecimento do envolvimento destes genes, bem como de outros mais directamente relacionados com a diabetes, dislipidemias e hipertensão seria bastante interessante no sentido de confirmar se as diversas vias e patologias referidas como associadas ao AVC são efectivamente factores de risco, tal como tem sido apontado (Kirshner, 2009).

Já no que respeita aos factores de risco não genéticos, estes são muito variados, abrangendo desde factores não modificáveis até ao estilo de vida adoptado pelo indivíduo.

Os factores de risco não genéticos e não modificáveis são geralmente a idade, o género e a etnia. Geralmente existe uma maior prevalência de AVCs em mulheres que em homens, enquanto os indivíduos de ascendência africana manifestam mais episódios que os caucasianos. A idade, por sua vez, assume um papel de destaque no risco de AVC. Após os 55 anos estima-se que a probabilidade de vir a sofrer da patologia duplique a cada dez anos (Brass, 1992; Lloyd-Jones *et al.*, 2009). Desta forma, justifica-se que os AVCs acima dos 65 anos sejam considerados de causa não genética, enquanto os que ocorrem em idades mais jovens estão relacionados com uma componente genética mais forte que potencia o seu desenvolvimento precoce. Quanto aos factores de risco não genéticos modificáveis, podem enumerar-se o consumo de álcool e tabaco, a alimentação, a ausência da prática de exercício físico, a terapia hormonal de substituição e o controlo de diversas patologias crónicas como a hipertensão, a diabetes, a hipercolesterolemia (Brass, 1992; Ringleb *et al.*, 2011; Warlow *et al.*, 2003). O consumo excessivo de álcool e de tabaco há muito que tem sido relacionado com o aumento do risco de doenças cardiovasculares. Enquanto o primeiro condicionará a alteração do perfil lipídico, o segundo potenciará o desenvolvimento de processos inflamatórios (Humphries & Morgan, 2004). A alimentação tem, por sua vez, um papel fundamental no decréscimo do risco de AVC, sempre que equilibrada, variada e adequada às necessidades do indivíduo. De facto, uma manutenção do peso corporal e, consequentemente, do índice de massa corporal, previne o desenvolvimento de diversas

patologias como a obesidade (e todas as complicações que dela advêm), hipertensão e diabetes, que constituem factores de risco para o AVC (Ringleb *et al.*, 2011). A prática de exercício físico está relacionada com uma melhoria da saúde cardiovascular, nomeadamente ao nível da redução da probabilidade de hipertensão, diabetes e doença cardíaca, e é igualmente um factor importante na redução do risco de AVC (Brass, 1992; Ringleb *et al.*, 2011). A terapia hormonal de substituição tem sido relacionada com o aumento do risco de AVC. Contudo, os mecanismos subjacentes necessitam de um estudo mais aprofundado (Brass, 1992). O controlo de patologias crónicas é fundamental na tentativa de redução do risco de AVC. Indivíduos que padeçam de hipertensão, diabetes e hipercolesterolemia, entre outras, são indivíduos de risco. Apesar disso, um controlo adequado dos níveis de tensão arterial, glucose e colesterol (nomeadamente através de medicação) poderá ajudar a aumentar a sua qualidade de vida e a prevenir os eventos cardiovasculares (Brass, 1992; Kirshner, 2009; Ringleb *et al.*, 2011).

O controlo dos factores de risco modificáveis é, então, a chave para uma prevenção efectiva do AVC em indivíduos com idade superior a 65 anos (já que em pacientes mais jovens as causas de AVC serão determinadas geneticamente não sendo tão influenciadas por factores modificáveis). Sempre que necessário, o acompanhamento médico de patologias crónicas deve ser efectuado da forma mais efectiva possível.

#### **1.1.6. Terapêutica**

A abordagem terapêutica de reperfusão numa fase aguda de um AVC isquémico tem-se revelado efectiva quando administrado plasminogénico activador tecidual (tPA) até 3h após o evento (Silverman *et al.*, 2011).

Um coágulo sanguíneo tem na sua constituição uma importante componente de fibrina. Digerida a fibrina, este torna-se passível de proteólise por outras enzimas, sendo rapidamente dissolvido. O tPA, protease de serina, desempenha então a sua função ao nível da clivagem do plasminogénico numa forma activa de plasmina, o qual digere a fibrina. Apesar de estudos no sentido do desenvolvimento de novas terapêuticas, a trombólise com tPa é, ainda, a única terapêutica que se tem revelado de facto efectiva (Adibhatla & Hatcher 2008; Silverman *et al.*, 2011).

Adicionalmente, outros tipos de terapêuticas têm sido aconselhados de forma continuada. A reabilitação física e a terapia da fala têm-se mostrado efectivas, assim como a administração de antitrombóticos de forma preventiva após o primeiro evento (Brass, 1992; Ringleb *et al.*, 2011).

#### **1.1.7. Epidemiologia**

Os AVCs constituem uma das principais causas de morte e incapacidade nos países desenvolvidos, sendo, logo a seguir à doença isquémica cardíaca, a segunda principal causa de morte devida a doenças cardiovasculares (Correia *et al.*, 2004; Guo *et al.*, 2010). Estima-se que existam cerca 1.1 milhões novos casos por ano na União Europeia, Suíça, Islândia e Noruega e que, nos mesmos países, aproximadamente 6 milhões de pessoas já tenham sofrido pelo menos um AVC (Truelsen *et al.*, 2005).



Quando comparadas regiões orientais e ocidentais da Europa, observam-se diferenças no que respeita à incidência e prevalência de AVCs, bem como na taxa de mortalidade associada a esta doença (Ringleb *et al.*, 2011). De acordo com Ringleb e colaboradores (2011), as divergências podem ser causadas pela desigual prevalência de factores de risco, como a hipertensão que se encontra largamente mais associada à Europa oriental e que, portanto, acabará por provocar nessa região AVCs mais severos. No entanto, considerando diferentes sub-regiões da Europa ocidental, existem igualmente variações pontuais nas mesmas taxas (Ringleb *et al.*, 2011).

Devido ao número de pacientes vítimas de AVC que permanecem incapacitados, os governos vêm-se obrigados a investir nos cuidados de saúde, quer no sector da reabilitação e cuidados continuados, quer em campanhas de prevenção, uma vez que se sabe que alguns estilos de vida (como, por exemplo, a prática de exercício físico) ajudam a diminuir o risco de ocorrência de AVC (Ekman, 2004; Hankey *et al.*, 2002; Hardie *et al.*, 2004). Para além dos custos já mencionados, existem ainda custos indirectos para o Estado, no que respeita a dias de trabalho perdidos (não só por parte dos próprios doentes, como também por parte dos seus familiares), e custos relativos a cuidados não-médicos (como transporte de doentes e adaptação de estruturas) (Andlin-Sobocki *et al.*, 2005; Ekman, 2004). Em 2004, os AVCs foram a segunda doença neurológica mais dispendiosa na Europa, levando a um gasto de cerca de 21 900 milhões de euros em aspectos com ela relacionados; em Portugal, foram gastos 853 milhões de euros, o que coloca o nosso país na 6ª posição do *ranking* de países com maior despesa associada a AVCs (Andlin-Sobocki *et al.*, 2005). No mesmo ano, Portugal teve uma das maiores taxas da Europa no que respeita à incidência e prevalência de AVC (considerando indivíduos com idade igual ou superior a 25 anos) (Truelsen *et al.*, 2005).

## **CAPÍTULO 2. SUSCEPTIBILIDADE AO AVC: A RELEVÂNCIA DO GENE *ADH1B***

## 2.1. INTRODUÇÃO

O consumo de bebidas alcoólicas pelo Homem não é uma prática recente, uma vez que existem evidências da produção de bebidas fermentadas já no tempo dos Faraós Egípcios e dos Sumérios da Mesopotâmia. Contudo, a discussão acerca dos benefícios e dos riscos desta prática continua a ser um tema actual, acerca do qual se têm desenvolvido estudos científicos, no sentido de perceber quais as vias fisiológicas afectadas pelo etanol e os consequentes efeitos na saúde (Kloner & Rezkalla, 2007).

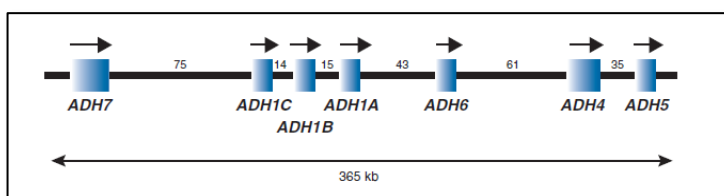
### 2.1.1. Metabolismo do álcool

O metabolismo do álcool no organismo humano é um processo complexo que ocorre predominantemente nos hepatócitos, através da actividade de duas famílias de proteínas: as álcool desidrogenases (ADHs) e as aldeído desidrogenases (ALDHs). Numa primeira etapa, o etanol é oxidado em acetaldeído através de ADHs no citosol e, posteriormente, este é oxidado em acetato e água por ALDHs nas mitocôndrias das mesmas células. Dada a toxicidade do acetaldeído, a quantidade deste metabolito que se acumula nos tecidos influencia as potenciais propriedades cardioprotectoras, bem como os já reconhecidos efeitos nocivos do consumo de álcool para a saúde. A referida acumulação de acetaldeído depende, entre outros factores, da quantidade de álcool ingerido e da relação entre a sua velocidade de síntese e degradação, a qual é condicionada pela funcionalidade das enzimas ADH e ALDH, respectivamente. Assim, é importante conhecer e compreender os diversos polimorfismos dos genes *ADH* e *ALDH*, que influenciam a taxa de metabolização do álcool e, desta forma, condicionam os efeitos destas bebidas na integridade do sistema cardiovascular (Husemoen *et al.*, 2008; Jelski & Szmikowski, 2008).

Conhecer os factores genéticos que determinam o metabolismo do etanol em cada indivíduo torna-se fundamental, quer para entender os hábitos de cada um, quer para tirar o melhor partido da prevenção que é possível fazer através desta via (Husemoen *et al.*, 2008).

#### Proteínas da família das álcool desidrogenases (ADHs)

Actualmente são conhecidas sete isoenzimas ADHs humanas, o maior número de ADHs identificado numa espécie. Cada isoenzima é codificada por um gene distinto de um *cluster* de genes localizado no braço longo do cromossoma 4 (4q) (Figura 2) (Edenberg, 2007). Ainda que todas as proteínas ADH humanas partilhem uma estrutura dimérica e a função enzimática de degradação do álcool, estas divergem entre si quando considerados outros parâmetros. Assim, é possível classificar estas proteínas em 5 classes (I-V), de acordo com as suas propriedades cinéticas e imunológicas, com a estrutura do DNA e sub-unidades proteicas codificadas (Cheung *et al.*, 1999).



**Figura 2** – Dimensão e posição relativas dos sete genes *ADH* humanos, ao longo do braço longo do cromossoma 4 (4q). As setas indicam o sentido em que ocorre a transcrição de cada gene, ainda que a região mais proximal ao centrómero seja a de *ADH5*. As distâncias entre genes estão indicadas em kilobases (kb) (Adaptado de Edenberg, 2007).

***ADH1B* e *ADH1C***

A generalidade das proteínas ADH são responsáveis pelo metabolismo do etanol e certos alelos nos respectivos genes condicionam a reacção de um indivíduo ao álcool. Existem duas ADHs em particular, a *ADH1B* e a *ADH1C*, ambas pertencentes à classe I, com um papel determinante na degradação do etanol em acetaldeído. Desta forma, tem-se verificado que os polimorfismos observados nos genes *ADH1B* e *ADH1C* modulam a actividade enzimática das proteínas codificadas, condicionando a taxa de conversão do álcool em acetaldeído e, consequentemente, a tolerância individual ao álcool e o risco do seu consumo para a saúde (Edenberg, 2007; Husemoen *et al.*, 2008; Yukawa *et al.*, 2012).

Observando os diversos haplotipos dos genes *ADH1B* e *ADH1C* (Tabela 2) verifica-se que o gene *ADH1B* tem um papel preponderante no que respeita à oxidação do etanol (Edenberg, 2007).

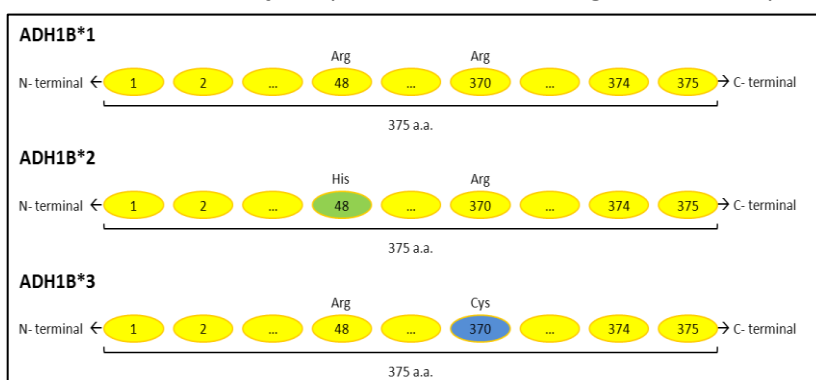
**Tabela 2** - Relação observada entre um determinado genótipo e o nível de metabolização de etanol (Edenberg *et al.*, 2007)

Genótipo	Nível de metabolização de etanol
<i>ADH1B</i> *1 + <i>ADH1C</i> *1	x
<i>ADH1B</i> *1 + <i>ADH1C</i> *2	< x (inferior)
<i>ADH1B</i> *2 + <i>ADH1C</i> *1	8 x (superior)
<i>ADH1B</i> *3 + <i>ADH1C</i> *1	2 x (superior)

x – valor de referência

O gene *ADH1B* é trialélico (*ADH1B*\*1, *ADH1B*\*2 e *ADH1B*\*3) e cada um dos alelos é responsável pela substituição de um aminoácido no produto proteico (Figura 3). Enquanto o alelo \*1 é considerado o alelo de referência, por ser o mais comum a nível mundial, estudos comparativos têm permitido concluir que os alelos \*2 (predominante em populações asiáticas) e \*3 (presente em populações de origem africana) regulam um aumento da capacidade oxidativa, o que se traduz numa mais rápida conversão de etanol em acetaldeído. Em termos comportamentais, indivíduos com os alelos \*2 e \*3 são menos susceptíveis a problemas de alcoolismo, por dois motivos: por um lado porque os níveis de etanol no organismo não permanecem elevados durante um grande intervalo de tempo; por outro, porque a produção de níveis elevados de acetaldeído, conduz os indivíduos a um estado de intoxicação que os dissuade do consumo excessivo de bebidas alcoólicas. A referida intoxicação pode sofrer um agravamento quando simultaneamente se encontram presentes determinadas isoformas de *ALDHs*, nomeadamente o alelo \*2 do gene *ALDH2*, relacionado com um défice severo na metabolização do acetaldeído (Chen *et al.*, 1999; Carr *et al.*, 2002; Edenberg, 2007).

Ainda que tenha sido possível relacionar ambos os alelos, \*2 e \*3, com a tolerância ao álcool, tem



**Figura 3** – Representação esquemática da proteína *ADH1B* codificada por cada um dos três alelos do gene *ADH1B* (Edenberg, 2007).

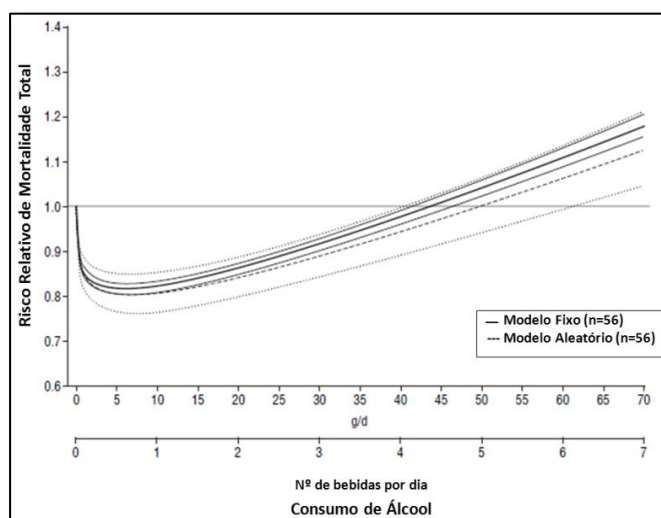
sido o alelo \*2 aquele que se tem revelado de maior interesse. Tem sido relativamente a este que se encontram os resultados com maior nível de significância nos estudos populacionais já realizados. Através do estudo do *Single Nucleotide Polymorphism* (SNP) rs1229984 (G/A), responsável pela substituição de uma arginina (alelo \*1 do gene *ADH1B*; alelo G do SNP rs1229984) por uma histidina (alelo \*2 do gene *ADH1B*; alelo A do SNP rs1229984) na posição 48 da proteína codificada (*Arg48His*), observou-se uma associação entre o alelo A do SNP e o consumo reduzido de bebidas alcoólicas. Assim, crê-se que o alelo A estará associado a um efeito protector em relação ao desenvolvimento de problemas decorrentes do alcoolismo (Macgregor *et al.*, 2009).

### 2.1.2. Relação entre o consumo de álcool e a doença cardiovascular (DCV)

Se tem sido possível associar o consumo excessivo de bebidas alcoólicas com o aumento da mortalidade, tem sido igualmente verificado que um consumo moderado está associado à sua redução (Figura 4) (DiCastelnuovo *et al.*, 2006).

#### Consumo excessivo vs moderado de álcool

O consumo de bebidas alcoólicas encontra-se frequentemente associado a factores socio-culturais e a comportamentos potenciados por polimorfismos genéticos envolvidos quer no metabolismo do álcool, quer em diversos tipos



**Figura 4** – Gráfico representativo da relação entre o consumo de álcool e a mortalidade (Kloner & Rezkalla, 2007).

de psicopatologias. No que respeita ao metabolismo do álcool, observou-se uma associação entre polimorfismos dos genes *ADH* e *ALDH* e certos padrões de consumo (Chen *et al.*, 1999; Macgregor *et al.*, 2009).

Tal como referido anteriormente, os alelos \*2 dos genes *ADH1B* e *ALDH2* (que codificam *Arg48His* na proteína *ADH1B* e *Glu504Lys* na *ALDH2*) estão associados, respectivamente, a uma aceleração na taxa de oxidação do etanol em acetaldeído e a um decréscimo da capacidade de eliminação deste através da sua conversão em acetato (Macgregor *et al.*, 2009). Sabendo que a acumulação de acetaldeído é tóxica, seria de esperar que, na presença dos alelos mencionados, os indivíduos consumissem menores quantidades de álcool e assim a susceptibilidade ao alcoolismo diminuísse. Contudo, não se verifica uma linearidade entre o genótipo dos indivíduos e os comportamentos no que respeita ao consumo de álcool (Chen *et al.*, 1999).

Estudos conduzidos em grupos de indivíduos com problemas de dependência do álcool mostram isso mesmo: indivíduos com as variantes genéticas protectoras contra o alcoolismo podem ainda assim consumir álcool em excesso e sofrer de forma mais severa as consequências directamente relacionados com essa prática (aumento do risco de doença coronária, arritmia ventricular e AVC) e as consequências relacionadas

com a acumulação de acetaldeído (interacção com o DNA através da indução de mutações e potenciação da formação de aductos) (Kloner & Rezkalla, 2007; Husemoen *et al.*, 2008; Yukawa *et al.*, 2012).

Relativamente aos indivíduos com os alelos de referência nos genes *ADH1B* e *ALDH2*, o seu comportamento admitirá maiores variações na taxa de consumo de álcool. Indivíduos que não manifestam qualquer tipo de reacção adversa ao consumo de bebidas alcoólicas consomem-nas, tendencialmente, em maior quantidade e, eventualmente, de forma excessiva. À semelhança do que acontece nos indivíduos portadores do alelo \*2, estes desenvolvem mais frequentemente patologias derivadas do abuso de álcool. No entanto, indivíduos com o alelo \*1 podem, por outro lado, adoptar o comportamento esperado por parte dos portadores do alelo \*2, abstendo-se do consumo de bebidas alcoólicas (motivados por outros aspectos que não os genéticos), ou consumindo álcool de forma moderada, o que tem sido relacionado com efeitos benéficos para o sistema cardiovascular (Chen *et al.*, 1999; DiCastelnuovo *et al.*, 2006; Drogan *et al.*, 2012).

De facto, existem evidências no sentido de um consumo moderado de bebidas alcoólicas ter efeitos benéficos para a saúde cardiovascular, nomeadamente na diminuição do risco de doença coronária. Estes benefícios serão mediados por uma optimização ao nível da função endotelial (maior vasodilatação), ao nível de processos inflamatórios (menor número de glóbulos brancos, decréscimo das concentrações de fibrinogénio e proteína C-reactiva, bem como um aumento das citocinas anti-inflamatórias), ao nível do perfil lipídico (nomeadamente pelo aumento dos níveis de lipoproteínas de alta densidade ou HDL) e ao nível da agregação plaquetária e da coagulação sanguínea. Igualmente importante, ainda que menos evidente, é o registo da diminuição da prevalência de hipertensão nestes indivíduos, provavelmente através da diminuição dos níveis de noradrenalina e vasopressina e do aumento dos níveis de péptido natriurético atrial (DiCastelnuovo *et al.*, 2006; Husemoen *et al.*, 2008; Imhof *et al.*, 2004; Kloner & Rezkalla, 2007; Rimm & Moats, 2007).

Relativamente à ocorrência de AVCs em particular, a relação entre o consumo de álcool e o risco associado não se encontra bem esclarecida. Contudo, Kloner e Rezkalla (2007) apontam para uma ligeira diminuição do risco face ao consumo ligeiro de álcool.

## 2.2. ENQUADRAMENTO E OBJECTIVOS

### 2.2.1. Enquadramento

Este trabalho tem como objectivo geral o estudo da associação entre um marcador genético funcional do gene *ADH1B*, o SNP rs1229984, e a susceptibilidade ao AVC e inclui-se num projecto desenvolvido em colaboração entre o grupo de Neurogenética e Saúde Mental (NSM), do Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis (DPS) do Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA), Lisboa, e o grupo de *Genetic Epidemiology*, do Departamento de Epidemiologia e Saúde Pública, da University College of London, Londres. O grupo *Genetic Epidemiology*, coordenado pelo Doutor Juan Pablo Casas, dedica-se ao estudo das implicações e aplicações da investigação em Genética na saúde, quer a nível individual, quer a nível de saúde pública e procura, através do seu trabalho, explorar três vertentes

principais: o potencial de estudos genéticos de larga-escala com vista à modelação do efeito de novas terapias; o potencial de estudos populacionais na compreensão da etiologia de doenças crónicas; o desenvolvimento e aplicação de métodos apropriados ao *design*, realização e análise de estudos populacionais com informação genómica incorporada, como forma de desenvolver ferramentas de predição de factores de risco e de inferência acerca da causalidade de diversas patologias. Neste âmbito, tem vindo a ser desenvolvido um trabalho centrado no SNP rs1229984, do gene *ADH1B*, uma vez que este SNP se tem revelado um marcador genético funcional associado a diversos fenótipos relacionados com a resposta de indivíduos de diversas origens ao álcool, a qual, por sua vez, influencia a função protectora do álcool sobre o desenvolvimento de doenças cardiovasculares (DCVs).

Com este estudo pretende confirmar-se a associação entre as variantes do SNP rs1229984 e os perfis fenotípicos relacionados com a tolerância ao álcool e investigar se este SNP pode ser utilizado como instrumento de predição em estudos de associação entre o álcool e DCVs.

### 2.2.2. Objectivos

O contributo dado pelo grupo NSM à equipa londrina é pertinente no âmbito do estudo da associação entre o SNP rs1229984 e um fenótipo de DCV, mais concretamente em termos de ocorrência de AVC.

Assim, pretende-se verificar se existe uma associação entre o SNP e a susceptibilidade ao AVC e se esta associação é mediada pelo consumo de álcool.

Após genotipagem individual do SNP na amostra e através de análise estatística, pretende-se poder estabelecer uma associação entre o *locus* e o consumo de álcool e entre o consumo de álcool e a susceptibilidade à ocorrência de um AVC.

## 2.3. METODOLOGIA

### 2.3.1. Caracterização da amostra

Os indivíduos participantes neste estudo foram recrutados a partir de duas amostras populacionais cujo material genético já se encontrava disponível no laboratório. Num total de 1002 indivíduos, 569 eram casos e 433 controlos. As características demográficas e clínicas da amostra encontram-se na Tabela 3.

### 2.3.2. Processamento da amostra

Amostras de DNA extraídas de sangue periférico encontravam-se já armazenadas no laboratório.

Recentemente, tendo sido necessário estudar novamente este material biológico, as amostras foram requantificadas, recorrendo ao espectrofotómetro NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE), por forma a verificar não só qual a concentração de cada amostra, como também a sua integridade, já que o equipamento tem a funcionalidade de calcular os rácios 260/230 e 260/280, indicativos da pureza do DNA. As amostras com valores dúbios dos rácios 260/230 e 260/280 foram corridas em gel de agarose 0.8% para confirmação da sua integridade.

**Tabela 3** – Caracterização da amostra populacional utilizada no estudo da relação entre o gene *ADH1B* e a susceptibilidade ao AVC

Característica	Controlos	Casos	P
<b>Idade e Género</b>			
Idade, média±DP (anos)	64.0±5.5	51.6±8.9	<10 <sup>-4*</sup>
Género (masculino), n/N (%)	201/433 (46.4)	362/569 (63.6)	<10 <sup>-4*</sup>
<b>Tipo de AVC</b>			
Isquémico, n/N (%)	-	451/569 (79.3)	-
Hemorragico, n/N (%)	-	107/569 (18.8)	-
Desconhecido, n/N (%)	-	11/569 (1.9)	-
<b>Factores de risco relacionados com a ocorrência de AVCs</b>			
Hipertensão (>85-140mmHg), n/N (%)	157/426 (36.9)	313/500 (62.6)	<10 <sup>-4†</sup>
Diabetes, n/N (%)	52/412 (12.6)	85/526 (16.2)	0.128 <sup>†</sup>
Hipercolesterolemia, n/N (%) (colesterol > 200mg/dl)	296/433 (68.3)	329/519 (63.4)	0.108 <sup>†</sup>
Consumo de tabaco, n/N (%)	120/423 (28.4)	257/557 (46.1)	<10 <sup>-4†</sup>
Consumo de álcool, n/N (%)~			<10 <sup>-4†</sup>
Não consome	254/407 (62.4)	242/558 (43.4)	
Consumo moderado	82/407 (20.1)	49/558 (8.8)	
Consumo excessivo	71/407 (17.4)	267/558 (47.8)	

DP – desvio padrão

~O consumo de álcool foi dividido em três categorias segundo a quantidade de álcool ingerida diariamente pelos indivíduos. Não consome: <1 bebida/dia; Consumo moderado: 1 bebida/dia para indivíduos do sexo feminino e 1 ou 2 bebidas por dia para indivíduos do sexo masculino; Consumo excessivo: >1 bebida/dia para indivíduos do sexo feminino e >2 bebidas por dia para indivíduos do sexo masculino (Rimm & Moats, 2007)

\* Teste Mann-Whitney; <sup>†</sup> Teste *Chi*-quadrado de Pearson

Seguidamente efectuaram-se diluições das amostras *stock*, uma primeira fase, para soluções intermédias (25ng/μL) e, posteriormente, para soluções de trabalho (10ng/μL). A existência de soluções intermédias é de grande importância, já que estas permitem a manipulação do DNA sem necessidade de recorrer ao *stock* repetidamente, o que minimiza risco de contaminação do mesmo e salvaguarda, assim, a integridade das amostras. Já as soluções de trabalho, por outro lado, permitem a utilização directa do DNA em técnicas de genotipagem, como o *TaqMan® Drug Metabolism Genotyping Assay*.

### 2.3.3. Genotipagem individual

O SNP rs1229984, do gene humano *ADH1B*, foi genotipado em 1002 indivíduos recorrendo a um *TaqMan® Drug Metabolism Genotyping Assay* (Applied Biosystems, Foster City, CA), lido num sistema de detecção de sequências ABI PRISM 7900HT (Applied Biosystems, Foster City, USA).

#### *TaqMan® Drug Metabolism Genotyping Assay*

Os *TaqMan® Drug Metabolism Genotyping Assay* são ensaios pré-desenhados, optimizados para a genotipagem de inserções, deleções, *Multi-Nucleotide Polymorphisms* (MNP) e SNPs. Existem cerca de 2000 ensaios diferentes disponíveis no mercado, desenhados em 220 genes envolvidos no transporte e metabolismo de drogas e validados através de genotipagem em cerca de 200 amostras humanas (Applied Biosystems, 2010).

No caso da genotipagem de SNPs, o procedimento a realizar passa por uma reacção de polimerização em cadeia (PCR) inicial para amplificação do fragmento que contém o SNP. As condições da mistura de reacção e



do programa utilizado encontram-se, respectivamente, descritas nas Tabela 4 e Tabela 5 (Applied Biosystems, 2010).

**Tabela 4** – Condições da mistura de reacção utilizada na execução do ensaio *TaqMan® Drug Metabolism Genotyping Assay* para a genotipagem do SNP rs1229984

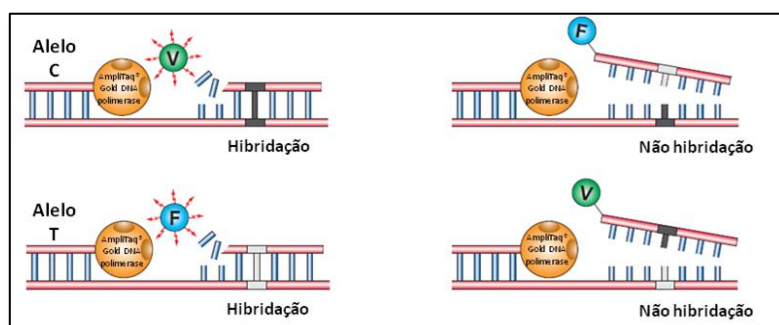
Reagentes	X1 reacção (µl)
<b>2X TaqMan Universal PCR Master Mix, No AmpErase UNG</b>	2.5
<b>20X Drug Metabolism Genotyping Assay Mix</b>	0.25
<b>Água DNase-free</b>	2.25
<b>Volume Total</b>	5.0

\* a mistura de reacção deve ser preparada ligeiramente em excesso por forma a compensar possíveis erros de pipetagem

**Tabela 5** – Programa utilizado na execução do ensaio *TaqMan® Drug Metabolism Genotyping Assay* para a genotipagem do SNP rs1229984

Etapas	Temperatura (°C)	Tempo	Nº de ciclos
<b>Activação da enzima AmpliTaq Gold - HOLD</b>	95	10'	1
<b>Desnaturação</b>	92	15"	50
<b>Annealing/Extensão</b>	60	1'30"	50

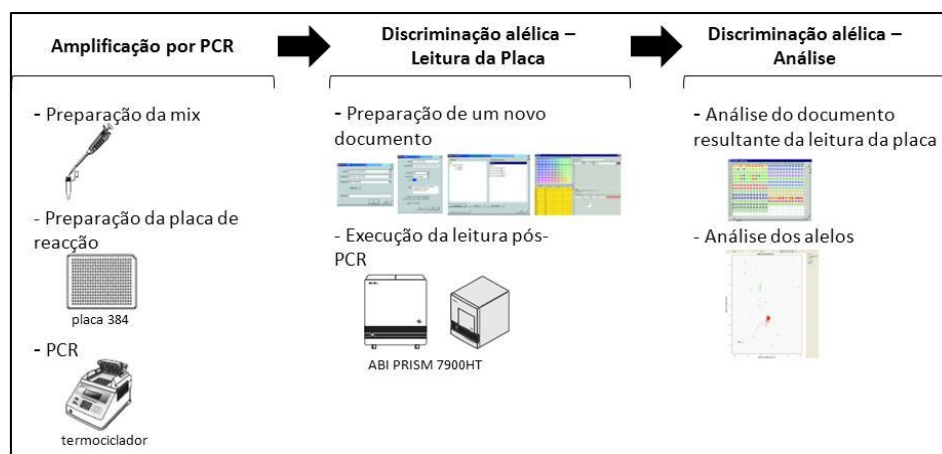
Esta PCR difere das PCRs convencionais na medida em que, para além de desoxirribonucleótidos trifosfatados (dNTPs) que complementam a cadeia de DNA do fragmento a amplificar, existem também sondas que hibridam especificamente na região do polimorfismo e às quais se encontra acoplado um fluoróforo, VIC ou 6-FAM, mediante o alelo presente (Figura 5) (Applied Biosystems, 2010).



**Figura 5** – Representação esquemática do PCR realizado no *TaqMan Drug Metabolism Genotyping Assay* aquando da genotipagem do SNP rs1229984. A sonda referente ao alelo C tinha acoplado o fluoróforo VIC, enquanto a do alelo T tinha o 6-FAM. (Adaptado de Applied Biosystems, 2010)

No caso do SNP rs1229984, polimorfismo G/A, as sondas do ensaio contêm dNTPs complementares, ou seja, C/T. O fluoróforo VIC está associado ao alelo C, enquanto o 6-FAM está acoplado ao T. Após a referida amplificação do fragmento contendo o SNP é feita uma análise da fluorescência, através do equipamento ABI PRISM 7900HT (Applied Biosystems, Foster City, USA) (Figura 6) (Applied Biosystems, 2010).

Mediante a intensidade da fluorescência detectada, assim as amostras são classificadas em homozigóticas (para cada alelo) ou heterozigóticas.



**Figura 6** – Representação esquemática das três fases da técnica de *TaqMan Drug Metabolism Genotyping Assay* (Adaptado de Applied Biosystems, 2010).

### Controlo de qualidade

O controlo de qualidade do ensaio foi efectuado através da utilização de indivíduos de genótipo conhecido (indivíduos HapMap) como controlos positivos; amostras duplicadas como controlos inter-ensaio; amostras de água, sem DNA, como controlos negativos; e verificação do Equilíbrio de Hardy-Weinberg (HWE) no grupo de controlos. A verificação do HWE permite inferir acerca da estabilidade da distribuição das frequências alélicas e genotípicas do SNP na população em geral entre gerações (daí calcular-se no grupo de controlos que são, precisamente representativos da população em geral). Sempre que os seus valores são  $>0.05$  existe evidência de uma estabilidade da distribuição das frequências alélicas e genotípicas do SNP e, no contexto de genotipagem, uma certeza de que não ocorreu nenhum erro no processo (Lunetta, 2008).

### 2.3.4. Análise estatística dos resultados

Uma vez que o estudo pretendia observar se uma possível associação entre os alelos e o risco de AVC seria mediada pelo consumo de álcool, após a genotipagem do SNP rs1229984 nos indivíduos seleccionados, realizou-se uma análise do tipo caso-controlo em que, através do cálculo e interpretação dos valores de *p-values* e *odds ratio* obtidos a partir de um teste *chi-quadrado* de Pearson, foi possível inferir acerca do papel específico de cada variável sobre o fenótipo observado.

A análise estatística efectuada foi realizada recorrendo aos softwares PLINK (v1.07) (Purcell *et al.*, 2007) e SPSS (v20.0) (IBM Corp., 2011).

## 2.4. RESULTADOS

O SNP rs1229984 foi genotipado em 1002 indivíduos, distribuídos por placas de 384 poços, nas quais se incluíram diversos controlos positivos, inter-ensaio e negativos. Os resultados são apresentados na Tabela 6.

O número final de indivíduos analisados foi 965, já que 37 dos 1002 indivíduos analisados não apresentaram informação para, pelo menos, um dos seguintes parâmetros em estudo: SNP ou consumo de álcool.

Os controlos positivos e inter-ensaio mostraram-se consistentes e os negativos não apresentaram qualquer anomalia. O valor do *p-value* associado ao Equilíbrio de Hardy-Weinberg (HWE) no grupo das amostras controlo revelou-se superior a 0.05 ( $P=0.262$ ), pelo que foi possível concluir que a genotipagem foi bem sucedida. A determinação do HWE não foi realizada incluindo o grupo de casos pois neste grupo um valor de HWE  $<0.05$  pode reflectir uma associação existente entre um genótipo e a susceptibilidade ao AVC e não necessariamente que tenha ocorrido qualquer erro na genotipagem.

Uma vez que o número de indivíduos com o genótipo AA é muito reduzido e, por isso, não seria possível analisá-lo estatisticamente de forma correcta, optou-se por associar os resultados dos genótipos AA e AG.

Note-se que o alelo A neste SNP é considerado o alelo \*2 do gene e o G o alelo \*1, *wild-type*.

**Tabela 6** – Resultados da genotipagem individual do SNP rs1229984

AVC	Consumo de álcool	Genótipo			Total
		GG	AG	AA	
Casos	Não consome	208	32	2	242
	Consumo moderado	43	6	0	49
	Consumo excessivo	242	23	2	267
	Total	493	61	4	558
Controlos	Não consome	214	37	3	254
	Consumo moderado	63	17	2	82
	Consumo excessivo	58	12	1	71
	Total	335	66	6	407
Total	Não consome	422	69	5	496
	Consumo moderado	106	23	2	131
	Consumo excessivo	300	35	3	338
	Total	828	127	10	965

Efectuou-se um teste *chi-quadrado* de Pearson, a fim de verificar a existência de associação entre as três variáveis consideradas, duas a duas. A

**Tabela 7** ilustra a análise preliminar de associação entre o genótipo e o AVC, enquanto as Tabela 8 e Tabela 9 mostram os resultados da análise preliminar de associação entre o genótipo e o consumo de álcool e o consumo de álcool e o AVC, respectivamente.

**Tabela 7** – Análise de associação entre o genótipo e o AVC

Genótipo		AVC		Total	P
		Casos	Controlos		
AA e AG	N	65	72	137	0.008
	%	11.6%	17.7%	14.3%	
GG	N	493	335	828	
	%	88.4%	82.3%	85.7%	
Total	N	558	407	965	
	%	100.0%	100.0%	100.0%	

**Tabela 8** – Análise de associação entre o genótipo e o consumo de álcool

Genótipo		Consumo de álcool			Total	P
		Não consome	Consumo moderado	Consumo excessivo		
AA e AG	N	74	25	38	137	0.074
	%	14.9%	19.1%	11.2%	14.2%	
GG	N	422	106	300	828	
	%	85.1%	80.9%	88.8%	85.8%	
Total	N	496	131	338	965	
	%	100.0%	100.0%	100.0%	100.0%	

**Tabela 9** – Análise de associação entre o consumo de álcool e o AVC

Consumo de álcool		AVC		Total	P
		Casos	Controlos		
Não consome	N	242	254	496	<10 <sup>-4</sup>
	%	43.4%	62.4%	51.4%	
Consumo moderado	N	49	82	131	
	%	8.8%	20.1%	13.6%	
Consumo excessivo	N	267	71	338	
	%	47.8%	17.4%	35.0%	
Total	N	558	407	965	
	%	100.0%	100.0%	100.0%	

Tal como se pode observar, as análises de associação genótipo-AVC e consumo de álcool-AVC revelaram resultados significativos, ou seja, admite-se que não existe independência entre as variáveis. Já no caso da análise de associação genótipo-consumo de álcool, os resultados não revelaram qualquer associação, embora exista uma tendência para um menor consumo excessivo de álcool no grupo de indivíduos portadores do alelo A e maior consumo excessivo no grupo de indivíduos GG. Contudo, seria interessante testar se existe alguma associação entre o genótipo e o consumo de álcool apenas no grupo de controlos. Uma vez que se sabe que o álcool é um factor de risco para o AVC, poder-se-ia colocar a hipótese da análise ser enviesada por um consumo excessivo de álcool por parte do grupo de casos, advindo a patologia precisamente desse comportamento, ainda que geneticamente houvesse uma predisposição para um consumo diferente do praticado.

Tal como se pode ver na Tabela 10, não há um resultado estatisticamente significativo em nenhum dos grupos (casos e controlos) quando se realiza a análise de associação proposta, pelo que não existe uma associação entre o genótipo neste SNP e o consumo de álcool.

**Tabela 10** - Análise de associação entre o genótipo e o consumo de álcool, considerando os grupos de casos e controles separadamente

AVC	Genótipo		Consumo de álcool			Total	P
			Não consome	Consumo moderado	Consumo excessivo		
Casos	AA e AG	N	34	6	25	65	0.256
		%	14.0%	12.2%	9.4%	11.6%	
	GG	N	208	43	242	493	
		%	86.0%	87.8%	90.6%	88.4%	
	Total	N	242	49	267	558	
		%	100.0%	100.0%	100.0%	100.0%	
Controlos	AA e AG	N	40	19	13	72	0.306
		%	15.7%	23.2%	18.3%	17.7%	
	GG	N	214	63	58	335	
		%	84.3%	76.8%	81.7%	82.3%	
	Total	N	254	82	71	407	
		%	100.0%	100.0%	100.0%	100.0%	

Relativamente às duas análises de associação cujos resultados se revelaram estatisticamente significativos, pormenorizou-se a análise.

Acerca da relação genótipo-AVC testou-se se, quando introduzida a variável consumo de álcool, esta exerceria um efeito modificador sobre a relação anteriormente observada.

Tal como ilustra a Tabela 11, no grupo de indivíduos que consomem álcool em excesso há uma associação entre o genótipo e o AVC ( $P=0.034$ ). Através da análise dos valores dos *odds ratio*, a presença do alelo A do SNP rs1229984 exerce um efeito protector sobre o AVC, uma vez que reduz para cerca de metade a probabilidade de um individuo portador desenvolver um AVC. Pelo contrário, na ausência do referido alelo, a probabilidade de um individuo que consome excesso de álcool vir a desenvolver esta patologia aumenta para cerca do quádruplo do grupo de referência (indivíduos sem o alelo A e que não consomem bebidas alcoólicas).

Relativamente à relação consumo de álcool-AVC testou-se, por outro lado, se, quando introduzida a variável genótipo, esta exerceria um efeito modificador sobre a relação anteriormente observada.

Tal como se pode observar na Tabela 12, o genótipo tem um efeito modificador sobre a associação consumo de álcool-AVC. Através da análise dos valores dos *odds ratio*, é possível verificar que existe uma probabilidade aumentada dos indivíduos que consomem álcool em quantidades excessivas desenvolverem AVCs quando têm genótipo GG, enquanto a mesma probabilidade decresce em relação ao valor de referência (não consumir álcool e não possuir o alelo A, considerado protector para a ocorrência de AVCs) para uma exposição moderada ao álcool. Assim, o consumo excessivo de álcool apresenta-se como um importante factor de risco para o AVC, como já tinha sido evidenciado na Tabela 8, reforçado pelo genótipo GG e protegido pelo alelo A.

O papel protector do alelo A sobre o AVC, tal como na análise anterior, é também aqui evidenciado. Efectivamente, no grupo de indivíduos portadores do referido alelo, não só o risco de AVC perante um consumo moderado de álcool decresce bastante mais que no grupo de indivíduos que não tem o alelo A, como o risco de AVC perante o consumo excessivo não aumenta tanto como no grupo que não possui o alelo protector. Assim, a influência exercida pelo consumo de álcool sobre o AVC é modificada pelo genótipo no SNP rs1229984 do gene *ADH1B*.

**Tabela 11** – Efeito modificador do consumo de álcool sobre a associação entre o genótipo e o AVC

Consumo de álcool	Genótipo		AVC		Total	P	OR
			Casos	Controlos			
Não consome	AA e AG	N	34	40	74	0.596	0.85
		%	14.0%	15.7%	14.9%		
	GG	N	208	214	422		1
		%	86.0%	84.3%	85.1%		
	Total	N	242	254	496		
		%	100.0%	100.0%	100.0%		
Consumo moderado	AA e AG	N	6	19	25	0.124	0.46
		%	12.2%	23.2%	19.1%		
	GG	N	43	63	106		0.68
		%	87.8%	76.8%	80.9%		
	Total	N	49	82	131		
		%	100.0%	100.0%	100.0%		
Consumo excessivo	AA e AG	N	25	13	38	0.034	0.46
		%	9.4%	18.3%	11.2%		
	GG	N	242	58	300		4.17
		%	90.6%	81.7%	88.8%		
	Total	N	267	71	338		
		%	100.0%	100.0%	100.0%		
Total	AA e AG	N	65	72	137	0.008	
		%	11.6%	17.7%	14.2%		
	GG	N	493	335	828		
		%	88.4%	82.3%	85.8%		
	Total	N	558	407	965		
		%	100.0%	100.0%	100.0%		

**Tabela 12** - Efeito modificador do SNP sobre a associação entre o consumo de álcool e o AVC

Genótipo	Consumo de álcool		AVC		Total	P	OR
			Casos	Controlos			
AA e AG	Não consome	N	34	40	74	0.005	0.85
		%	52.3%	55.6%	54.0%		
	Consumo moderado	N	6	19	25		0.32
		%	9.2%	26.4%	18.2%		
	Consumo excessivo	N	25	13	38		1.92
		%	38.5%	18.1%	27.7%		
	Total	N	65	72	137		
		%	100.0%	100.0%	100.0%		
GG	Não consome	N	208	214	422	<10 <sup>-4</sup>	1
		%	42.2%	63.9%	51.0%		
	Consumo moderado	N	43	63	106		0.68
		%	8.7%	18.8%	12.8%		
	Consumo em excesso	N	242	58	300		4.17
		%	49.1%	17.3%	36.2%		
	Total	N	493	335	828		
		%	100.0%	100.0%	100.0%		
Total	Não consome	N	242	254	496	<10 <sup>-4</sup>	
		%	43.4%	62.4%	51.4%		
	Consumo moderado	N	49	82	131		
		%	8.8%	20.1%	13.6%		
	Consumo em excesso	N	267	71	338		
		%	47.8%	17.4%	35.0%		
	Total	N	558	407	965		
		%	100.0%	100.0%	100.0%		

## 2.5. DISCUSSÃO

Ainda que a vertente sociocultural tenha um papel muito determinante no que respeita ao consumo de bebidas alcoólicas, sabe-se actualmente que esta não é o único factor por detrás de tal prática. Diversos estudos genéticos já demonstraram que existem polimorfismos ao nível dos genes que codificam as proteínas ADH e ALDH envolvidas no metabolismo do etanol, e que esses polimorfismos condicionam a reacção dos indivíduos a esta substância (Chen *et al.*, 1999; Macgregor *et al.*, 2009). É o caso do SNP rs1229984, localizado no gene *ADH1B*, que determina a inserção de uma arginina (alelo \*1 – *wild-type*) ou de uma histidina (alelo \*2) na posição 48 da proteína, mediante o alelo G ou A, respectivamente, se encontre presente. De acordo com as implicações funcionais do polimorfismo do referido SNP, que condiciona a tolerância ao álcool, seria de esperar que indivíduos homozigóticos para o alelo G não manifestassem qualquer efeito adverso ao álcool, enquanto os portadores do alelo A manifestassem uma intolerância considerável, traduzida comportamentalmente numa redução do seu consumo (Macgregor *et al.*, 2009). Contudo, no âmbito deste estudo, a relação entre o SNP rs1229984 e o consumo de álcool não é

óbvia, não tendo sido possível replicar as afirmações de Macgregor e colaboradores (2009). Quando realizada uma análise de associação entre o genótipo e o consumo de álcool, os resultados apresentaram-se não significativos ( $P>0.05$ ), quer considerando a amostra total, quer analisando separadamente os grupos de casos e controlos. Ainda assim, este resultado poderá não significar necessariamente que a relação entre o genótipo e a tolerância ao consumo de álcool seja falsa, mas sim que possa ter ocorrido um enviesamento neste estudo em particular, dadas as características da amostra:

- Por um lado, seria necessário possuir dados mais detalhados relativamente aos hábitos de consumo de álcool dos indivíduos, o que não aconteceu pelo facto dos indivíduos não terem sido recrutados especificamente para um estudo com estas exigências. A maior lacuna nos detalhes da amostra é a impossibilidade de distinguir os indivíduos que não consomem álcool motivados por factores socioculturais (motivos pessoais, religiosos, entre outros) daqueles que não consomem álcool por de facto apresentarem algum tipo de intolerância ao mesmo determinada geneticamente. Idealmente, os indivíduos que não consumissem álcool motivados por factores não genéticos, deveriam ter sido excluídos da amostra. Este factor poderá ser então o que provoca o maior enviesamento da análise. Em relação aos indivíduos que dizem consumir bebidas alcoólicas, desconhecem-se também eventuais efeitos adversos que se possam manifestar, decorrentes de uma intolerância determinada geneticamente, e que possam estar a interferir com a análise realizada.

- Por outro lado, o número de indivíduos analisados deveria ter sido maior, tendo em conta a prevalência do alelo menos frequente (A) na população CEU (<1%). Segundo Wang e colaboradores (2005), sempre que um alelo ligado à susceptibilidade a um fenótipo (neste caso à tolerância ao álcool) tem uma frequência alélica <1% na população estudada, o número de indivíduos estudados deveria ser aproximadamente 20 000 (nos quais a distribuição dos fenótipos – neste caso, os hábitos de consumo de bebidas alcoólicas – deveria ser equilibrada).

Para além das limitações encontradas relacionadas com a amostra populacional, existem ainda outros factores que podem modificar a associação entre o SNP rs1229984 e o consumo de álcool. Segundo Edenberg (2007), será possível que entre indivíduos haja uma expressão distinta de diversos genes, incluindo do *ADH1B*. Desta forma e tal como já foi observado em estudos anteriores, a tolerância ao álcool poderá ser distinta entre indivíduos com o mesmo genótipo em *loci* do gene *ADH1B*, nomeadamente na presença do alelo \*2 do gene *ALDH2*, por exemplo. Tendo em conta este factor, e ainda que se associe o alelo A do SNP rs1229984 a uma maior intolerância ao álcool, é difícil definir se geralmente a referida intolerância é total ou parcial.

Seria, portanto, interessante repetir esta análise, por forma a consolidar a potencial relação entre o SNP rs1229984 e o consumo de álcool apresentado pelos indivíduos.

Os padrões de consumo de bebidas alcoólicas têm sido associados a distintos níveis de risco para a saúde cardiovascular (Kloner & Rezkalla, 2007). Desta forma, o estudo da relação entre o SNP rs1229984, do gene *ADH1B*, e o fenótipo de DVC ganha relevância, que aumenta quando se verifica que o consumo de álcool é efectivamente modificador do efeito do genótipo sobre AVC.



Assim, realizou-se uma análise estratificada em que se observou a relação entre o referido SNP e a susceptibilidade ao AVC tendo em conta o padrão de consumo de bebidas alcoólicas. Os resultados do estudo mostram-se significativos apenas no grupo de indivíduos que consome álcool excessivamente. Ainda assim, é possível tirar conclusões que vão de encontro a resultados de estudos anteriores que apontam o álcool como factor de risco para o AVC e o alelo A do SNP em estudo como factor de protecção. (Ringleb *et al.*, 2011).

Uma interpretação dos valores dos *odds ratio* permite afirmar que, perante um consumo excessivo de bebidas alcoólicas, os indivíduos não portadores do alelo protector têm um risco de AVC quatro vezes superior ao grupo de referência (indivíduos que não consomem álcool nem possuem o alelo protector); por outro lado, os indivíduos portadores do alelo A têm o risco de sofrer um AVC reduzido para cerca de metade. Conclui-se que o alelo A do SNP desempenha efectivamente um papel protector sobre o risco de AVC, apesar do consumo excessivo de álcool. Contudo, estes resultados necessitam futuras réplicas, pois não foi possível observar qualquer tendência significativa nos grupos de indivíduos que não consomem álcool ou consomem doses moderadas (ainda que, empiricamente se observe uma tendência para um papel protector do alelo A, já que na sua presença, os valores de *odds ratio* tendem a ser sempre os mais baixos). Relativamente ao grupo de indivíduos que consome bebidas alcoólicas de forma moderada, teria sido particularmente interessante observar uma relação entre o genótipo e o fenótipo, uma vez que, segundo diversos autores, o consumo de álcool em doses moderadas é benéfico para a saúde cardiovascular, o que poderia traduzir uma menor prevalência de AVCs (DiCastelnuovo *et al.*, 2006; Husemoen *et al.*, 2008; Imhof *et al.*, 2004; Kloner & Rezkalla, 2007; Rimm & Moats, 2007).

Ainda que não fosse o objectivo principal deste estudo e, por isso, não tenha sido possível determinar a relevância do consumo de álcool sobre todas as variáveis que se pensa que este influenciará, foi possível analisar a relação entre o consumo de álcool e a susceptibilidade ao AVC em geral. Também esta análise de associação se revelou estatisticamente significativa. Na tentativa de observar se a relação entre o consumo de álcool e de AVC estariam a sofrer um efeito modificador por parte do SNP, através da presença do alelo considerado protector, realizou-se uma análise estratificada. Para ambos os grupos de indivíduos portadores e não portadores do alelo protector foi possível observar resultados estatisticamente significativos.

Numa primeira fase, a partir da observação dos valores dos *odds ratio*, é possível verificar que independentemente do genótipo, os indivíduos que consomem álcool exageradamente manifestam sempre uma tendência no sentido de um risco aumentado de sofrer AVC; por outro lado, verifica-se igualmente uma tendência para o consumo moderado de álcool ser protector, uma vez que os indivíduos que consomem álcool de forma moderada tendem a ter um risco diminuído. Desta forma, é possível confirmar que, mediante a quantidade de álcool consumida, assim esta substância pode ser um importante factor de risco ou um factor protector para o AVC. Os resultados obtidos vão de encontro ao que já havia sido descrito por Ringleb e colaboradores (2011), que afirmam que um consumo excessivo de álcool aumenta o risco de AVC quer isquémico, quer hemorrágico, relativamente a indivíduos que não consomem bebidas alcoólicas nas

mesmas doses elevadas, e vão também de acordo às afirmações de Kloner e Rezkalla (2007) que sugerem uma ligeira diminuição do risco de AVC face ao consumo ligeiro de álcool.

Os resultados tornam-se ainda mais interessantes quando analisados tendo em conta a presença ou ausência do alelo protector. De facto, enquanto indivíduos que não possuem o alelo A e consomem álcool moderadamente vêem o seu risco de AVC diminuído para 0.62, os indivíduos que possuem o alelo A vêem o seu risco diminuído para 0.32. Já quando se considera um consumo excessivo de álcool, os indivíduos portadores de A têm um risco de AVC de cerca do dobro dos indivíduos referência (não portadores do alelo A e não consumidores de álcool), enquanto os que não possuem o alelo A vêem o risco de AVC aumentado para cerca do quádruplo.

Assim, novamente se verifica a tendência da presença do alelo A do SNP rs1229984 ser um factor protector.

Porém, é importante não esquecer que todas as análises de associações entre genótipo e fenótipos neste estudo são de elevada complexidade, uma vez que as características são poligénicas. Assim, quanto mais abrangentes e completos a análise e os dados, melhores e mais concretos resultados se poderão esperar.

Também ao nível das implicações do consumo moderado de álcool sobre diversas vias fisiológicas, seria interessante proceder a estudos funcionais e fisiológicos. Em particular e dada a incidência cada vez mais alta de hipercolesterolemia e hipertensão, seria importante determinar até que ponto e segundo que mecanismos o consumo de álcool pode ser preventivo nestes casos, tal como sugerem diversos autores (DiCastelnuovo *et al.*, 2006; Husemoen *et al.*, 2008; Imhof *et al.*, 2004; Kloner & Rezkalla, 2007; Rimm & Moats, 2007).

Por outro lado, seria também importante continuar a proceder à educação das populações no sentido de alertar para os malefícios do consumo excessivo de álcool. Segundo a Organização Mundial de Saúde é este o segundo maior factor de risco para diversos tipos de patologias a nível mundial (WHO, 2011).

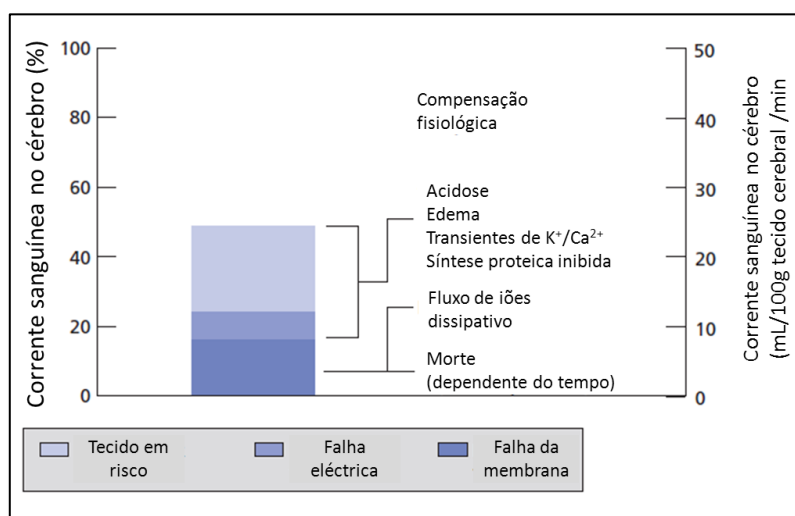
## **CAPÍTULO 3. IDENTIFICAÇÃO DE VARIANTES GENÉTICAS QUE INFLUENCIAM A RECUPERAÇÃO APÓS UM AVC ISQUÊMICO**

### 3.1. INTRODUÇÃO

Os AVCs isquémicos, sendo os mais comuns e com melhor prognóstico, são aqueles de despertam um maior interesse quando se estudam os factores que influenciam a recuperação após um evento.

A irrigação sanguínea ao cérebro em situações normais é de aproximadamente 50mL de sangue por cada 100g de tecido cerebral por minuto. Contudo, na presença de uma oclusão que provoque um AVC isquémico, esta irrigação diminui e a função neuronal é afectada. Se inicialmente o volume de sangue que irriga determinada região cerebral decai para cerca de metade, o que provoca a perda reversível das propriedades eléctricas dos neurónios, quanto maior o intervalo de tempo que decorrer entre o evento e a obtenção de cuidados de saúde por parte de um paciente, mais decairá o volume de sangue que irriga a região e maior o carácter irreversível das lesões. Uma irrigação inferior a 10mL de sangue por cada 100g de tecido cerebral por minuto desencadeia falhas no metabolismo mitocondrial aeróbio e no metabolismo da glucose, o que conduz a um estado de acidose. Consequentemente, a homeostase celular iónica dependente de energia falha, o que desequilibra as concentrações de sódio, potássio e cálcio nas células e culmina na morte neuronal (Figura 7) (Davenport & Dennis, 2000).

Tendo um forte impacto sobre a funcionalidade dos neurónios, não é difícil compreender que os AVCs isquémicos deixem os pacientes incapacitados.



**Figura 7** – Limiar da corrente sanguínea cerebral para a disfunção celular e morte (Adaptado de Warlow *et al.* 2007).

Após um AVC, a recuperação dos pacientes pode ser muito heterogénea: em cerca de 50-70% dos casos, existe uma recuperação total da independência funcional; por outro lado, cerca de 15-30% dos pacientes fica permanentemente incapacitado; e em 20% dos casos três meses após o AVC os pacientes ainda requerem cuidados institucionais (Manso *et al.*, 2010). Segundo estudos de recuperação realizados a médio e longo-prazo, alguns autores propõem que cerca de 33% dos pacientes que sobrevive 5 anos após um primeiro AVC permanecem incapacitados, assim como 50% dos pacientes que sobrevivem mais de 10 anos (Hankey *et al.* 2002; Hardie *et al.*, 2004). É igualmente sugerido que, quanto mais severo o AVC, maior o grau

de incapacidade, o que leva a uma probabilidade diminuída de recuperação e aumentada de dependência a longo-prazo (Hankey *et al.* 2002). Consoante a idade do paciente, também a probabilidade de recuperar a independência funcional até 6 meses após o AVC é distinta: para idades inferiores a 65 anos existe uma probabilidade de 64%, ao passo que para pessoas com idade igual ou superior a 80 anos a mesma probabilidade decresce para os 20% (Hankey *et al.*, 2002). Deste modo, existirá um risco de morbilidade e morte aumentados e uma probabilidade de melhoria diminuída para pacientes mais idosos (Hankey *et al.*, 2002). Em termos de recorrência, um estudo a longo-prazo conduzido por Hardie e colaboradores (2004), mostrou que o risco de ocorrência de um novo AVC decorridos um ou dez anos sobre o primeiro evento seria de 4% e 40%, respectivamente. Já a taxa de AVCs recorrentes que se revelam fatais ao fim de 30 dias corresponde a 40% (sensivelmente o dobro da probabilidade de um primeiro AVC ser fatal nos 30 dias seguintes à sua ocorrência) (Hardie *et al.*, 2004).

### 3.1.1. Escalas de classificação da recuperação

Quando um indivíduo sofre um AVC isquémico, uma das primeiras questões que se impõe tem a ver com o prognóstico. Assim, foram desenvolvidas várias escalas que permitem medir especificamente quer os défices neuronal e motor do indivíduo na fase aguda (*National Institutes of Health Stroke Scale – NIHSS*), por forma a decidir qual a melhor terapêutica a aplicar, quer a evolução da recuperação (*Modified Rankin Scale – mRS* – e *Barthel index – BI*), permitindo inferir acerca da reabilitação a médio e longo prazo, quer ainda o impacto que o sucedido teve sobre o paciente (*Stroke Impact Scale – SIS*) (Kasner, 2006).

A mRS (

Tabela 13) é a escala que mais frequentemente se utiliza em estudos de recuperação de AVCs, por apresentar uma informação mais completa acerca do impacto que o AVC teve no estilo de vida do paciente e na sua capacidade para realizar as actividades diárias ou para manter uma participação activa num contexto social (Kasner, 2006; Wilson *et al.*, 2002). Geralmente os indivíduos são classificados no momento do diagnóstico de AVC e após um determinado período de recuperação. O momento reconhecido como o mais apropriado para inferir acerca do nível da recuperação é 3 meses após o evento. Contudo, observa-se que uma classificação a 10 dias após o AVC confere uma boa estimativa da recuperação aos 3 meses, o que permite calcular aproximadamente qual o período de recuperação e respectivos recursos financeiros necessários (Ovbiagele & Saver, 2010).

### 3.1.2. Factores envolvidos na recuperação

A recuperação de um paciente após um AVC isquémico depende de diversos factores e, muitas vezes, da sua interacção (Warlow *et al.*, 2003). Por um lado, existem factores não genéticos inalteráveis que podem influenciar o prognóstico, como a idade, o género, doenças crónicas e a severidade do AVC (Brass, 1992; Lloyd-Jones *et al.*, 2009; Ringleb *et al.*, 2011; Warlow *et al.*, 2003); por outro, existem ainda factores modificáveis que são de extrema importância e que condicionam um maior ou menor sucesso da recuperação: é fundamental, por exemplo, que o intervalo de tempo decorrido até à obtenção de cuidados

**Tabela 13 – Modified Rankin Scale (mRS) (adaptado de Kasner, 2006)**

mRS	Sintomas
0	Sem sintomatologia
1	Sem incapacidades significativas, à excepção de sintomas de AVC (apto a realizar as actividades e deveres usuais)
2	Incapacidade ligeira (incapaz de executar a totalidade das actividades anteriores mas capaz de cuidar de assuntos pessoais sem assistência)
3	Incapacidade moderada (exige alguma assistência, mas é capaz de andar autonomamente)
4	Incapacidade moderadamente grave (incapaz de andar autonomamente e incapaz de atender às necessidades fisiológicas sem assistência)
5	Incapacidade grave (acamado, incontinente, requer cuidados de enfermagem constantes)
6	Morte

médicos seja o mais curto possível (Ingall, 2004) e que, ao longo do processo de recuperação, o plano de reabilitação (através de fisioterapia e/ou terapia da fala e/ou terapia ocupacional) seja adequado e ajustado aos progressos do doente (Brass, 1992).

Relativamente à componente genética, existem ainda evidências que a apontam como determinante na recuperação dos pacientes.

As primeiras observações que levantaram esta hipótese advieram de estudos familiares. Lisabeth e colaboradores (2005) observaram uma relação estatisticamente significativa entre uma pior recuperação funcional e a existência de antecedentes familiares do paciente com a mesma doença. Os primeiros estudos no sentido de identificar variantes genéticas envolvidas na recuperação após um AVC isquémico foram realizados em ratinhos por Atochin e colaboradores (2010) que identificaram uma relação entre a proteína guanilato ciclase e a preservação da corrente sanguínea em situações de isquémia. Mais tarde, Maguire e colaboradores (2011) mostraram já na espécie humana, que variantes do gene *COX-2* estão associadas a uma melhor recuperação funcional, enquanto variantes do gene *GP1IIa* estão associadas a uma pior recuperação, medida segundo as escalas anteriormente descritas. Os genes candidatos que foram analisados nos diversos estudos de susceptibilidade ao AVC foram seleccionados pelo seu envolvimento em processos de dano e remodelação neurovascular (*MMP-2* e *MMP-9*), em vias metabólicas reguladoras do stress oxidativo no cérebro (*MTHFR* e *GSTO-1*) e em vias inflamatórias (*IL6*, *MPO*, *IL1B*, *TNF* e *PTGS2*). Destes, os genes *MMP-2* e *IL6* revelaram-se associados à recuperação; porém, os mecanismos afectados pelos polimorfismos nestes genes requerem ainda uma investigação mais aprofundada (Manso *et al.*, 2010; Manso *et al.*, 2011; Peddareddygar *et al.*, 2009).

Tal como se pode constatar, há ainda um longo caminho a percorrer no sentido de identificar variantes genéticas envolvidas na recuperação após um AVC isquémico e no sentido de compreender alterações por elas condicionadas nas vias fisiológicas afectadas pelo AVC. A abordagem preponderante neste tipo de estudos tem sido a do estudo de genes candidatos. Porém, existem outras abordagens cujos resultados é importante analisar, como é o caso dos *Genome Wide Association Studies* (GWAS) que têm um grande

potencial no que respeita à identificação de novos SNPs relacionados com uma determinada via (Pearson *et al.*, 2007).

### 3.1.3. Prevenção secundária

Outra das questões que se impõem quando um indivíduo sofre um AVC tem a ver com a probabilidade de reincidência (Warlow *et al.*, 2003). Esta encontra-se intimamente ligada aos factores de risco que actuam sobre um indivíduo (idade, hipertensão sistólica, embolismo sistémico prévio, entre outros), bem como à possível medicação e estilo de vida praticados (Ingall, 2004).

Ao controlo de factores modificáveis que afectam um indivíduo que já sofreu um AVC e que podem estar na origem da reincidência do evento, dá-se o nome de prevenção secundária (sendo a primária a prevenção realizada antes do registo de qualquer episódio de AVC). Desta fazem parte, entre outros: o controlo regular da tensão arterial, da glucose e do colesterol; a toma correcta de medicação prescrita pelo médico, seja para a diabetes, para prevenir a formação de coágulos – antitrombóticos, ou para qualquer outra finalidade; a descontinuação do consumo de tabaco e bebidas alcoólicas; a prática de exercício físico; uma dieta equilibrada; um controlo do índice de massa corporal, que deve permanecer dentro dos parâmetros adequados (Ringleb *et al.*, 2011; Warlow *et al.*, 2003).

Acções de sensibilização da população neste sentido seriam muito úteis e efectivas no combate a uma das maiores causas de morte e morbilidade a nível mundial.

## 3.2. ENQUADRAMENTO E OBJECTIVOS

### 3.2.1. Enquadramento

Este projecto, cujo objectivo geral é a identificação variantes genéticas envolvidas na recuperação após um AVC isquémico, constitui a continuação de um trabalho previamente iniciado com o mesmo objectivo: o projecto PTDC/SAU-GMG/64426/2006, financiado pela Fundação para a Ciência e Tecnologia (FCT).

#### Resumo do trabalho efectuado previamente

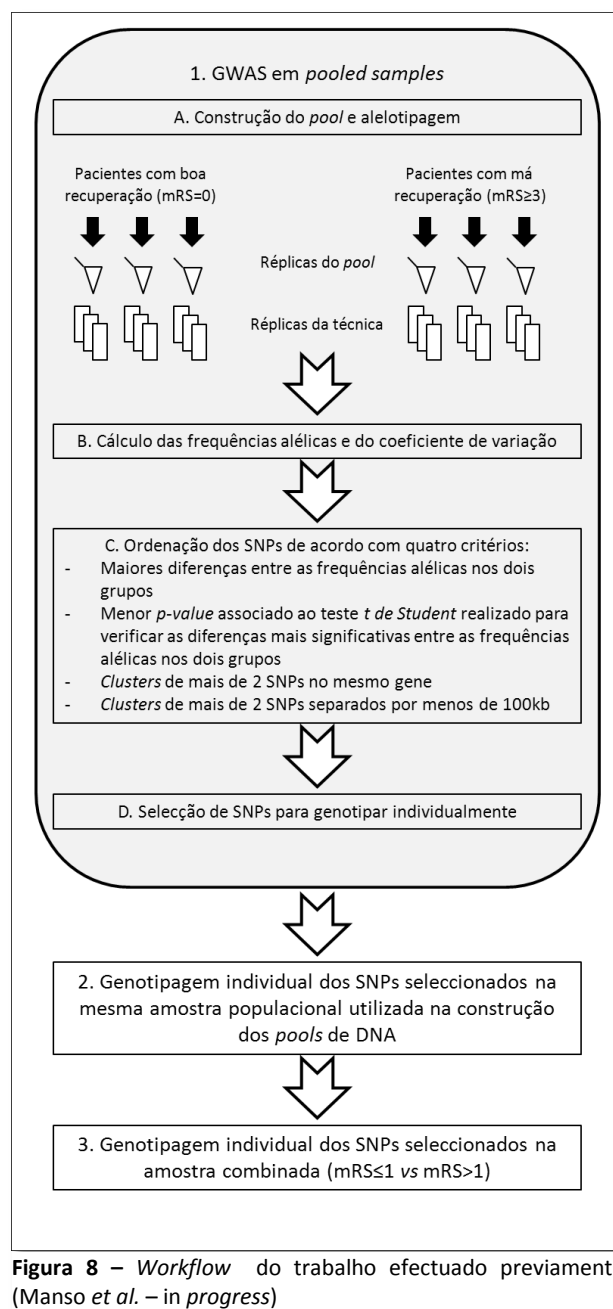
Numa primeira etapa do projecto, foi efectuado um rastreio genómico (*Genome-Wide Association Study* – GWAS) em grupos de amostras de DNA (*pooled samples*). O GWAS é uma estratégia de pesquisa de variantes genéticas associadas a uma determinada via ou doença complexa, que não pressupõe nenhuma hipótese prévia (*hypothesis free*), isto é, rastreia todo o genoma em busca de *loci* envolvidos num determinado processo biológico, analisando indiscriminadamente locais que empiricamente possam estar ou não associados a esse processo. No entanto, é uma técnica bastante dispendiosa. Com vista a colmatar esta limitação recorre-se à análise de *pooled samples*, que são amostras constituídas por DNA de diversos indivíduos que partilham um mesmo fenótipo face à característica que se esteja a analisar, em vez de se rastrear cada indivíduo separadamente (Pearson *et al.*, 2007; Macgregor, 2007). No caso deste trabalho, os pacientes foram subdivididos, segundo a escala mRS, em dois grupos com valores de recuperação extremos: com muito boa recuperação (mRS =0) (N=87) e com má recuperação (mRS ≥3) (N=100). Para cada um dos

grupos, foram realizados três *pools* semelhantes de DNA (réplicas de formação de *pool*) e em cada uma das réplicas foi realizada uma alelotipagem em triplicado de 262 264 SNPs, recorrendo ao 250K Affymetrix GeneChip® Mapping Assay – Nsp I (Affymetrix, Santa Clara, CA, USA). A alelotipagem em triplicado é uma réplica da técnica, que permite determinar a variância resultante da alelotipagem nos diversos *pools*, funcionando assim como controlo de qualidade, nomeadamente na identificação de SNPs cujos resultados reflectam dificuldades técnicas e que, por isso, devam ser excluídos de uma análise futura (Pearson *et al.*, 2007; Macgregor, 2007). Em suma, foi feito um total de 9 réplicas para cada grupo (*pool*). (Figura 8-1.A) Através de uma análise estatística preliminar, para controlo de qualidade da técnica e identificação de SNPs associados à recuperação, foram calculados (1) o valor médio da frequência alélica de cada SNP em cada grupo e (2) respectivo coeficiente de variação (CV) (indicador da dispersão de resultados referentes a um mesmo SNP nas várias réplicas) (Figura 8-1.B), (3) a diferença entre os valores médios obtidos para cada grupo e (4) a significância da referida diferença (através de um teste estatístico *t*-student). Com base nestes dados, todos os SNPs com CV superior a 10% foram excluídos, na tentativa de eliminar falsos positivos, e posteriormente adoptaram-se quatro critérios que permitiram seleccionar os SNPs mais interessantes para genotipar (Figura 8-1.C e 1.D):

1) Ordenaram-se os SNPs por ordem decrescente das maiores diferenças entre as frequências alélicas dos dois grupos (boa e má recuperação) e admitiram-se os primeiros 1000 SNPs (*top 1000*). Destes, e com base no gráfico da distribuição das referidas diferenças, seleccionaram-se os primeiros 46 SNPs localizados na região de maior acentuação da curva (Figura Suplementar 1) (Abraham *et al.*, 2008).

2) Ordenaram-se os SNPs por ordem crescente do *p*-value associado ao teste *t* de Student realizado para verificar a significância das diferenças entre as frequências alélicas nos dois grupos e admitiram-se os primeiros 1000 SNPs (*top 1000 – t-test*). Destes, e novamente com base no gráfico da distribuição dos referidos *p*-values, seleccionaram-se 34 SNPs localizados mais próximo da origem dos eixos (Figura Suplementar 2) (Diergaarde *et al.*, 2010).

3 e 4) De entre os SNPs considerados no conjunto



**Figura 8 – Workflow** do trabalho efectuado previamente (Manso *et al.* – in progress)



dos *top 1000* e *top 1000 – t-test* seleccionaram-se ainda todos os que se encontravam em *clusters* de 3 ou mais SNPs separados por menos de 100kb (14 SNPs) ou localizados no mesmo gene (15 SNPs) (Abraham *et al.*, 2008; Baum *et al.*, 2008).

Dos 109 SNPs identificados, 9 foram seleccionados por mais de um critério. Um total de 100 SNPs foi então sujeito a validação, tendo cada SNP sido genotipado individualmente na amostra utilizada para constituição das *pooled samples* (N=187) (Figura 8-2). Após este procedimento excluíram-se 12 SNPs da análise por terem falhado o controlo de qualidade e prosseguiu-se com a análise dos restantes 88. Foi então possível validar 36 SNPs ( $P<0.05$ ), isto é, foi possível identificar através de uma regressão logística, 36 SNPs potencialmente associados ao padrão de recuperação após um AVC. Estes SNPs foram genotipados adicionalmente em 227 indivíduos com mRS=1 e mRS=2, não incluídos nas *pooled samples* (Figura 8-3). O objectivo de alargar a amostra foi o de verificar se os resultados se mantinham quando não se testavam apenas grupos extremos de recuperação, considerando o *cut-off* de boa/má recuperação entre o nível 1 e 2 da mRS (boa recuperação = mRS≤1 e má recuperação = mRS>1), já que clinicamente este é um valor mais adequado. A esta amostra (N=414) chamou-se amostra combinada. O resultado da genotipagem individual dos SNPs validados na amostra combinada permitiu concluir que 21 SNPs validados não tinham associação com o padrão de recuperação após um AVC, enquanto 15 SNPs, por outro lado, se mantinham consistentes com a análise prévia ( $P<0.05$ ). Contudo, ao incluir na análise factores que possam ser de confundimento, co-variáveis associadas ao padrão de recuperação após um AVC isquémico, apenas 6 SNPs mantiveram o seu *p-value* ao nível da significância. Estes resultados sugerem 2 genes, *BBS9* e *GLIS3*, associados ao processo de recuperação.

No presente trabalho pretendeu-se realizar uma análise mais abrangente. Os GWAS utilizando *pooled samples* têm associada uma grande variabilidade e, se valores elevados de CV podem incluir falsos positivos, o valor de 10%, pode, por outro lado, ser demasiado restritivo e excluir muitos falsos negativos. Adoptando este *cut-off*, excluiu-se da análise um número muito elevado de SNPs. Assim, seria interessante alargar o valor de CV (para 15%) e proceder a uma nova análise por forma a verificar se os SNPs agora incluídos seriam ou não falsos positivos e, consequentemente, qual o melhor *cut-off* a adoptar para o CV no contexto de GWAS recorrendo a *pooled samples*.

### 3.2.2. Objectivos

O objectivo do presente trabalho é a identificação de novas variantes genéticas (SNPs e, eventualmente, genes) potencialmente envolvidas na recuperação após um AVC, com CV associado (após o GWAS) entre 10% e 15%. O valor *cut-off* de 15% pretende aumentar a amostra de SNPs a estudar e determinar se este *cut-off* aumenta significativamente a frequência de falsos positivos.

### 3.3. METODOLOGIA

#### 3.3.1. Caracterização da amostra

Na sequência do trabalho previamente realizado para identificação de novas variantes genéticas envolvidas na recuperação após um AVC, os indivíduos envolvidos nesta fase do trabalho foram os mesmos que já tinham sido utilizados anteriormente.

A amostra global, denominada amostra combinada, integrou um total de 414 indivíduos (nos quais se incluíram os 187 integrantes dos *pools*) de idade inferior a 65 anos e que registavam a ocorrência de pelo menos um AVC isquémico (confirmado medicamente) até à data. A classificação da sua recuperação foi feita de acordo com a mRS e o *cut-off* para classificar uma boa recuperação vs uma má recuperação foi estabelecido entre os níveis 1 e 2 da mRS. As características demográficas e clínicas da amostra encontram-se na Tabela 14.

**Tabela 14** – Caracterização da amostra combinada utilizada no estudo de variantes genéticas associadas à recuperação após um AVC

Característica	Boa recuperação (mRS≤1)	Má recuperação (mRS>1)	P
<b>Idade e Género</b>			
Idade, média±DP (anos)	50.7±9.0	52.1±8.9	0.125*
Género (masculino), n/N (%)	148/230 (64.3)	116/184 (63.0)	0.784*
<b>Historial clínico, n/N (%)</b>			
Hipertensão	127/198 (64.1)	77/163 (47.2)	0.001 <sup>†</sup>
Diabetes	34/217 (15.7)	35/173 (20.2)	0.241 <sup>†</sup>
Doença cardíaca	33/218 (15.1)	33/175 (18.9)	0.327 <sup>†</sup>
<b>Características do AVC, n/N (%)</b>			
Afasia	43/217 (19.8)	56/173 (32.4)	0.005 <sup>†</sup>
Inatenção	9/223 (4.0)	14/165 (8.5)	0.067 <sup>†</sup>
Disfagia	13/227 (5.7)	19/175 (10.9)	0.060 <sup>†</sup>
Incontinência urinária	2/227 (0.9)	12/175 (6.9)	0.001 <sup>†</sup>
Paresia	168/227 (74.0)	165/184 (89.7)	<10 <sup>-4†</sup>
Alterações de consciência	8/230 (3.5)	31/181 (17.1)	<10 <sup>-4†</sup>
Complicações médicas	10/223 (4.5)	43/171 (25.2)	<10 <sup>-4†</sup>
Complicações neurológicas	12/228 (5.3)	19/183 (10.4)	0.051 <sup>†</sup>

DP – desvio padrão

\* Teste Mann-Whitney; <sup>†</sup> Teste Chi-quadrado de Pearson

#### 3.3.2. Processamento da amostra

Uma vez que as amostras para o presente trabalho já tinham sido utilizadas anteriormente, o DNA já se encontrava extraído, quantificado e diluído para as concentrações de 25ng/μl e 10ng/μl. A existência de soluções intermédias, de 25ng/μl, é de grande importância, já que estas, para além de permitirem a manipulação do DNA sem necessidade de recorrer ao *stock* repetidamente (o que minimiza o risco de contaminação do mesmo, salvaguardando a integridade das amostras), permitem também a utilização directa do DNA em técnicas de biologia molecular como o PCR. Já as soluções de trabalho, por outro lado, permitem a utilização directa do DNA em técnicas de genotipagem, como o Sequenom iPLEX Gold e o TaqMan® SNP Pre-Designed Genotyping Assay.

### 3.3.3. Selecção dos SNPs a analisar

A partir dos dados obtidos na alelotipagem das *pooled samples*, calculou-se o coeficiente de variação (CV) associado a cada SNP e excluíram-se todos aqueles com CV >15%. De seguida ordenaram-se os restantes SNPs por ordem decrescente das diferenças entre as frequências alélicas nos dois grupos e excluíram-se todos aqueles que se encontravam abaixo da posição 1000, obtendo-se o *top 1000*. Adoptaram-se então três critérios distintos para identificar os SNPs mais interessantes a genotipar individualmente:

1) Pré-seleccionaram-se os SNPs com melhor posição no *top 1000*, isto é, com as maiores diferenças entre as frequências alélicas nos dois grupos. Para definir o número de SNPs a seleccionar para genotipagem individual, analisou-se a curva representativa das referidas diferenças e estabeleceu-se o *cut-off* num ponto do gráfico em que a acentuação da curva visivelmente mudava (Figura 9). Foi então possível identificar 39 SNPs.

2) Pré-seleccionaram-se os *clusters* de 3 ou mais SNPs separados entre si por menos de 100kb. Através deste critério foi possível identificar 15 SNPs.

3) Pré-seleccionaram-se os *clusters* de 3 ou mais SNPs presentes no mesmo gene. Este parâmetro permitiu identificar 31 SNPs.

Tendo em conta que um mesmo SNP pode ser seleccionado por mais de um critério, a construção da lista de SNPs mais relevantes para genotipar individualmente incluiu 75 SNPs potencialmente interessantes. Contudo, excluíram-se ainda alguns SNPs de acordo com critérios muito bem definidos:

→ Não foram analisados os 3 SNPs que já tinham sido previamente analisados no trabalho antecedente com o mesmo objectivo.

→ Verificou-se o desequilíbrio de *linkage* (LD) ( $r^2=1.0$ ) entre os restantes SNPs recorrendo à base de dados do projecto HapMap (Consortium, 2003) visualizada no software Haploview (v4.2). Dos grupos de SNPs em LD, seleccionou-se apenas um para genotipar. Nesta fase, reduziu-se o número de SNPs a genotipar de 72 para 59.

→ Finalmente, observou-se que em dois *clusters* em que se haviam seleccionado múltiplos SNPs a posição de alguns destes no *top 1000* era muito elevada (>381). Uma vez que era esperado que os resultados mais significativos se encontrassem em SNPs muito melhor posicionados, optou-se por analisar apenas um SNP de cada um desses *clusters*, por forma a inferir acerca do interesse de genotipar os restantes.

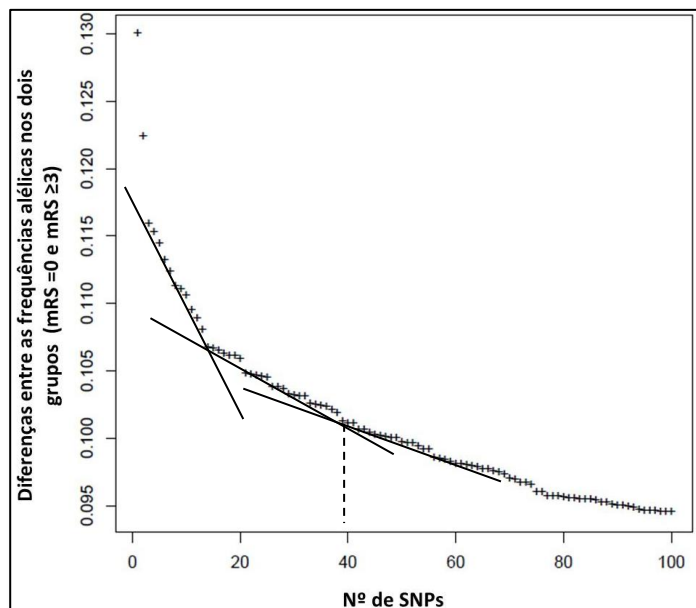


Figura 9 – Gráfico representativo das maiores diferenças entre as frequências alélicas nos dois grupos (mRS=0 e mRS≥3)

O número final de SNPs seleccionados para genotipagem individual foi de 52 e a caracterização dos mesmos encontra-se na Tabela 15.

Ao contrário do trabalho prévio, em que o resultado do teste *t de Student* realizado para verificar quais as diferenças mais significativas entre as frequências alélicas nos dois grupos também foi critério de selecção de SNPs, neste trabalho este critério não foi tido em conta, já que os SNPs que seriam seleccionados coincidiam com os que já tinham sido seleccionados anteriormente segundo o mesmo método.

**Tabela 15** – Caracterização dos SNPs genotipados individualmente

SNP	Posição top 1000	Cr	Posição	Alelos	Diferenças entre as frequências alélicas nos dois grupos mRS=0 e mRS≥3	Coefficiente de Variação no grupo mRS=0	Coefficiente de Variação no grupo mRS≥3	Distância entre SNPs <100kb?	Gene	Região	Critério(s) de selecção dos SNPs	Obs.
rs7514276	818	1	217031796	G/T	0.0794	12.0104	10.4272	SIM	LOC643717	intrão	cluster <100kb	
rs17048803	557	1	217130725	C/T	0.0823	13.3668	11.3232				cluster <100kb	
rs687417	38	1	217170628	A/C	0.1018	12.0651	8.5283				cluster <100kb; top 1000	
rs651539	16	1	234904056	T/C	0.1064	8.9496	14.9071				top 1000	
rs4660095	18	1	239706677	T/C	0.1060	12.0558	13.5562				top 1000	
rs10514631	7	2	19767453	T/C	0.1123	14.8245	14.1934				top 1000	
rs2699164	17	2	36335706	C/T	0.1062	12.9822	12.9832				top 1000	
rs17024145	19	2	39768185	A/G	0.1060	10.8990	8.0077		TMEM178	intrão	top 1000	
rs12711674	35	2	124670233	A/C	0.1023	6.8279	11.6570		CNTP5	intrão	top 1000	
rs1562646	77	2	186956727	A/G	0.0956	9.7900	13.7289	SIM			cluster <100kb	* <sup>1</sup>
rs2975788	22	2	241220531	G/A	0.1046	12.9923	8.9604				top 1000	
rs615329	26	3	4102195	A/G	0.1038	11.2813	12.8374				top 1000	
rs17235018	20	3	7187512	G/C	0.1058	11.7104	13.0721		GRM7	intrão	top 1000	
rs10049174	5	3	55116477	C/T	0.1144	8.4683	14.6874				top 1000	
rs11706616	3	3	175626076	T/C	0.1158	11.0096	13.2184				top 1000	
rs4916566	4	3	198083321	A/G	0.1152	13.7535	10.2751		SENP5	intrão	top 1000	
rs1479249	43	4	11852542	C/T	0.1005	10.6790	13.7837	SIM			cluster <100kb	* <sup>2</sup>
rs11933940	23	4	37238765	G/A	0.1045	13.4778	14.2428		C4orf19	intrão	top 1000	
rs4864056	21	4	132827716	T/C	0.1047	12.9021	14.9698				top 1000	
rs16887451	1	5	17954332	A/C	0.1299	11.2021	9.1959				top 1000	
rs11747001	8	5	132440198	A/G	0.1112	9.9982	12.4026		HSPA4	intrão	top 1000	
rs2876137	15	6	8347264	A/G	0.1066	14.1823	13.8708				top 1000	
rs9350278	210	6	20853694	T/A	0.0892	9.8179	12.2167	SIM	CDKAL1	intrão	cluster <100kb; cluster gene	+
rs4710950	744	6	20911078	A/T	0.0801	5.8950	8.0548			intrão	cluster <100kb; cluster gene	
rs10253978	566	7	33521197	A/G	0.0821	12.0976	11.3009		BBS9	intrão	cluster gene	~ <sup>1</sup>
rs2617091	978	8	4438281	T/G	0.0780	8.6394	9.1754		CSMD1	intrão	cluster gene	●
rs9314584	760	8	5758935	C/G	0.0800	12.3998	10.9097	SIM			cluster <100kb	
rs17075289	114	8	5791863	G/T	0.0936	10.6141	13.2414				cluster <100kb	
rs2703313	298	8	5805958	A/C	0.0866	12.2239	14.9188				cluster <100kb	
rs13258732	40	8	104179751	T/G	0.1011	9.8542	14.6923				top 1000	
rs13249775	48	8	109494368	C/T	0.1000	10.3934	12.4086	SIM	LOC100128549	intrão	cluster <100kb	* <sup>3</sup>
rs7860786	39	9	94285294	C/T	0.1012	12.1650	11.3956		CENPP	intrão	top 1000	
rs10903686	800	10	2190219	G/A	0.0796	9.8906	4.3162				cluster <100kb	
rs1243411	12	10	9009808	G/T	0.1088	8.1103	10.5682				top 1000	
rs11001208	14	10	76363007	T/C	0.1066	11.1615	14.0857		MYST4	intrão	top 1000	

rs10830772	29	11	91158549	A/G	0.1032	9.4835	13.5583			top 1000
rs11175781	24	12	64167286	T/G	0.1045	10.9686	12.3804			top 1000
rs7980667	331	12	101641189	A/G	0.0859	10.8015	11.3370	SIM		cluster <100kb ~ <sup>2</sup>
rs11849729	32	14	45564779	C/T	0.1030	13.8234	10.9145			top 1000
rs11159233	10	14	76295162	C/A	0.1105	8.5697	11.2134			top 1000
rs2044068	13	14	87597185	T/C	0.1079	8.4786	14.7010			top 1000
rs16949386	31	15	93049222	A/G	0.1030	12.5134	14.6889			top 1000
rs4888652	37	16	76195008	G/A	0.1021	11.1835	14.0439			top 1000
rs7207469	11	17	532225	T/G	0.1094	11.4596	14.4336	VPS53	intrão	top 1000
rs6506157	34	18	4030656	A/G	0.1024	11.9073	8.9843			top 1000
rs1144046	30	18	48282979	G/A	0.1031	10.9166	14.9189	DCC	intrão	top 1000
rs6039208	149	20	8536762	A/G	0.0915	14.0032	14.4253			intrão cluster gene
rs6140752	33	20	8770308	T/G	0.1025	8.4972	14.9972			intrão cluster gene; top 1000
rs6086654	848	20	8786376	C/T	0.0791	12.6805	14.9972			intrão cluster gene
rs6012664	28	20	47486019	C/T	0.1036	12.3897	14.6340	KCNB1	intrão	top 1000
rs5960190	9	X	56651979	T/C	0.1110	12.7858	11.2054			top 1000
rs17218233	25	X	125146041	A/G	0.1044	13.1238	13.5906			top 1000

Os polimorfismos indicados referem-se ao sentido *forward*.

O primeiro alelo indicado corresponde ao alelo ancestral.

\*<sup>1</sup> – os restantes SNPs do *cluster* são os SNPs rs4667081 (posição 297 no *top 1000*), rs12475405 (posição 780 no *top 1000*), rs4666705 (posição 645 no *top 1000*), rs4144784 (posição 474 no *top 1000*) e rs12693453 (posição 182 no *top 1000*) e estão em LD com este.

\*<sup>2</sup> – os restantes SNPs do *cluster* são os SNPs rs1479280 (posição 330 no *top 1000*), rs2124141 (posição 933 no *top 1000*) e rs6448875 (posição 118 no *top 1000*) e estão em LD com este.

\*<sup>3</sup> – os restantes SNPs do *cluster* são os SNPs rs7816458 (posição 333 no *top 1000*) e rs13271967 (posição 641 no *top 1000*) e estão em LD com este.

+ – o terceiro SNP do *cluster* é o SNP rs9295481 (posição 719 no *top 1000*) e está em LD com este.

~<sup>1</sup> – os restantes SNPs do *cluster* são os SNPs rs10486524 (posição 27 no *top 1000*) e rs10273634 (posição 574 no *top 1000*) que estão em LD entre si. O SNP rs10273634 já havia sido genotipado anteriormente.

~<sup>2</sup> – os restantes SNPs do *cluster* são os SNPs rs7131780 (posição 132 no *top 1000*) e rs10778191 (posição 2 no *top 1000*) que estão em LD entre si. O SNP rs7131780 já havia sido genotipado anteriormente.

● – os restantes SNPs do *cluster* foram excluídos devido à sua posição geral e excessivamente elevada no *top 1000* e ao facto de não ser possível determinar a existência de LD entre si (rs584541 - posição 955 no *top 1000*, rs3849836 - posição 577 no *top 1000*, rs10503246 - posição 797 no *top 1000*, rs10112705 - posição 706 no *top 1000* e rs2028237 - posição 381 no *top 1000*)

O SNP na posição 6 do *top 1000* não foi considerado por não cumprir o requisito de ser bialélico e o da posição 36, rs12830815 já havia sido genotipado anteriormente.

### 3.3.4. Genotipagem individual

Os SNPs seleccionados foram genotipados através de três técnicas distintas: RFLP (*Restriction Fragment Length Polymorphism*), Sequenom iPLEX Gold e TaqMan® *SNP Pre-Designed Genotyping Assay*. A Tabela 16 ilustra quais os SNPs genotipados com cada técnica.

A análise do SNP rs6012664 foi substituída pela análise do SNP rs742758, por uma questão relativa à técnica de genotipagem. Uma vez que a genotipagem com Sequenom iPLEX Gold analisa conjuntos de SNPs (*plex*s) e o primeiro SNP referido não pôde ser incluído em nenhum *plex* por incompatibilidade com os restantes, tendo em conta que o LD entre os SNPs era  $r^2=1.0$ , procedeu-se à substituição pelo segundo SNP, que já pôde então ser incluído num *plex* e genotipado com Sequenom iPLEX Gold.

**Tabela 16** – Resumo dos SNPs genotipados recorrendo a cada uma das três técnicas utilizadas para o efeito

SNP	Cr	Posição	Alelos	Técnica utilizada para a genotipagem individual
rs7514276	1	217031796	G/T	Sequenom iPLEX Gold
rs17048803	1	217130725	C/T	Sequenom iPLEX Gold
rs687417	1	217170628	A/C	Sequenom iPLEX Gold
rs651539	1	234904056	T/C	RFLP
rs4660095	1	239706677	T/C	Sequenom iPLEX Gold
rs10514631	2	19767453	T/C	Sequenom iPLEX Gold
rs2699164	2	36335706	C/T	Sequenom iPLEX Gold
rs17024145	2	39768185	A/G	Sequenom iPLEX Gold
rs12711674	2	124670233	A/C	Sequenom iPLEX Gold
rs1562646	2	186956727	A/G	Sequenom iPLEX Gold
rs2975788	2	241220531	G/A	RFLP
rs615329	3	4102195	A/G	Sequenom iPLEX Gold
rs17235018	3	7187512	G/C	RFLP
rs10049174	3	55116477	C/T	Sequenom iPLEX Gold
rs11706616	3	175626076	T/C	Sequenom iPLEX Gold
rs4916566	3	198083321	A/G	Sequenom iPLEX Gold
rs1479249	4	11852542	C/T	Sequenom iPLEX Gold
rs11933940	4	37238765	G/A	Sequenom iPLEX Gold
rs4864056	4	132827716	T/C	RFLP
rs16887451	5	17954332	A/C	<i>TaqMan® SNP Pre-Designed Genotyping Assay</i>
rs11747001	5	132440198	A/G	Sequenom iPLEX Gold
rs2876137	6	8347264	A/G	Sequenom iPLEX Gold
rs9350278	6	20853694	T/A	Sequenom iPLEX Gold
rs4710950	6	20911078	A/T	Sequenom iPLEX Gold
rs10253978	7	33521197	A/G	Sequenom iPLEX Gold
rs2617091	8	4438281	T/G	Sequenom iPLEX Gold
rs9314584	8	5758935	C/G	Sequenom iPLEX Gold
rs17075289	8	5791863	G/T	Sequenom iPLEX Gold
rs2703313	8	5805958	A/C	Sequenom iPLEX Gold
rs13258732	8	104179751	T/G	Sequenom iPLEX Gold
rs13249775	8	109494368	C/T	Sequenom iPLEX Gold
rs7860786	9	94285294	C/T	Sequenom iPLEX Gold
rs10903686	10	2190219	G/A	Sequenom iPLEX Gold
rs1243411	10	9009808	G/T	Sequenom iPLEX Gold
rs11001208	10	76363007	T/C	Sequenom iPLEX Gold
rs10830772	11	91158549	A/G	<i>TaqMan® SNP Pre-Designed Genotyping Assay</i>
rs11175781	12	64167286	T/G	Sequenom iPLEX Gold
rs7980667	12	101641189	A/G	Sequenom iPLEX Gold
rs11849729	14	45564779	C/T	Sequenom iPLEX Gold
rs11159233	14	76295162	C/A	Sequenom iPLEX Gold
rs2044068	14	87597185	T/C	<i>TaqMan® SNP Pre-Designed Genotyping Assay</i>
rs16949386	15	93049222	A/G	Sequenom iPLEX Gold

rs4888652	16	76195008	G/A	Sequenom iPLEX Gold
rs7207469	17	532225	T/G	Sequenom iPLEX Gold
rs6506157	18	4030656	A/G	Sequenom iPLEX Gold
rs1144046	18	48282979	G/A	Sequenom iPLEX Gold
rs6039208	20	8536762	A/G	Sequenom iPLEX Gold
rs6140752	20	8770308	T/G	Sequenom iPLEX Gold
rs6086654	20	8786376	C/T	Sequenom iPLEX Gold
rs742758	20	47495219	A/C	Sequenom iPLEX Gold
rs5960190	X	56651979	T/C	Sequenom iPLEX Gold
rs17218233	X	125146041	A/G	Sequenom iPLEX Gold

#### 3.3.4.1. RFLP (*restriction fragment length polymorphism*)

A técnica RFLP é uma das técnicas mais antigas utilizadas para a detecção manual de polimorfismos, nomeadamente de SNPs. Consiste na amplificação por PCR da sequência a analisar, seguida de digestão enzimática, na qual uma endonuclease específica cliva o fragmento obtido na PCR exactamente no *locus* de interesse. Posteriormente através da análise dos tamanhos dos fragmentos obtidos é possível inferir acerca do genótipo presente.

#### PCR

A *Polimerase Chain Reaction* (PCR) é uma técnica de Biologia Molecular utilizada para amplificar fragmentos de DNA, *in vitro* e de forma exponencial. À semelhança do que ocorre *in vivo*, a PCR mimetiza a replicação do DNA no ciclo celular e conta com 3 etapas cíclicas: desnaturação, *annealing* e extensão (Roche, 2006).

O número de ciclos a realizar é determinado pela quantidade de DNA que é pretendido ou necessário obter. Para que a reacção seja bem-sucedida acrescentam-se ainda, normalmente, 2 etapas não cíclicas à reacção: uma desnaturação inicial e uma extensão final.

Para que a PCR seja bem-sucedida, é necessário juntar ao DNA *template* uma mistura de reacção composta por reagentes com concentrações bem definidas. No caso deste trabalho, a composição desta mistura de reacção encontra-se ilustrada na Tabela 17. Os *primers* específicos para cada fragmento a amplificar (desenhados recorrendo ao programa *online* Primer3 (v.0.4.0) disponível em [frodo.wi.mit.edu](http://frodo.wi.mit.edu) (Rozen & Skaletsky, 2000)), por sua vez, constam na Tabela suplementar 1.

Os intervalos de tempo, temperaturas e número de ciclos utilizados nas PCRs dos SNPs a analisar por RFLP encontram-se descritos na Tabela 18. Relativamente às temperaturas de *annealing*, é importante referir que, ainda que estas se baseiem na sequência nucleotídica dos *primers* e respectiva temperatura de *melting*, nem sempre estes valores de temperatura coincidem com os da referida temperatura de *melting*, sendo por vezes necessário proceder a uma optimização desta condição da técnica.

**Tabela 17** – Quantidades e concentrações dos reagentes no volume final de reação de PCR para genotipagem com RFLP

Reagentes	volume (μl)	[final]
Solução tampão (Promega)	2	1X
MgCl <sub>2</sub> (Promega)	0.6	1.5mM
dNTPs* (Promega)	0.2	10 μM (cada dNTP)
Primer F	1	1μM
Primer R	1	1μM
DNA Polimerase (Promega)	0.05	0.25U
Água	4.15	n.a.
DNA	1	2.5ng/μl
<b>Total</b>	<b>10</b>	

\* foi feita uma mistura de dNTPs na qual a concentração de cada um era 2mM; n.a. – não se aplica.

**Tabela 18** – Programas utilizados nas PCRs dos SNPs a analisar por RFLP

	Intervalo de tempo	Temperatura (°C)	rs651539	rs2975788	rs17235018	rs4864056	Nº ciclos
Desnaturação inicial	5'	95°C					
Desnaturação	1'	95°C					
Annealing	30"		66°C	61°C	61°C	59°C	35
Extensão	1'	72°C					
Extensão final	5'	72°C					
Pausa	∞	4°C					

### Digestão enzimática

Após proceder à amplificação do fragmento de interesse por PCR, é adicionada ao produto uma enzima de restrição ou endonuclease. Estas enzimas, reconhecem sequências de DNA de 4-6 pares de bases e clivam-nas numa posição nucleotídica específica e conhecida, originando diversos fragmentos de comprimento igualmente conhecido.

Para identificar as enzimas mais adequadas a cada polimorfismo que se pretenda analisar, é possível recorrer a uma ferramenta *online*, o SNP-rflping (disponível em <http://bio.kuas.edu.tw/snp-rflping/>) (Chang *et al.*, 2006) que, mediante a inserção da sequência contendo o SNP de interesse, indica quais as enzimas que a clivam no local de interesse e na presença de qual dos alelos.

No caso dos quatro SNPs genotipados por RFLP, as enzimas utilizadas, suas características e condições de utilização encontram-se na Tabela 19.

### Análise dos fragmentos obtidos

Após a digestão enzimática com endonucleases específicas para cada caso, os fragmentos obtidos foram analisados através de electroforese em gel de agarose (Lonza).

No caso dos 4 SNPs analisados, o padrão de bandas esperado e correspondente genótipo encontram-se esquematizados na Figura Suplementar 3 – Representação esquemática dos locais de corte das enzimas de restrição utilizadas nas sequências amplificadas por PCR e padrão de fragmentos associados a cada genótipo

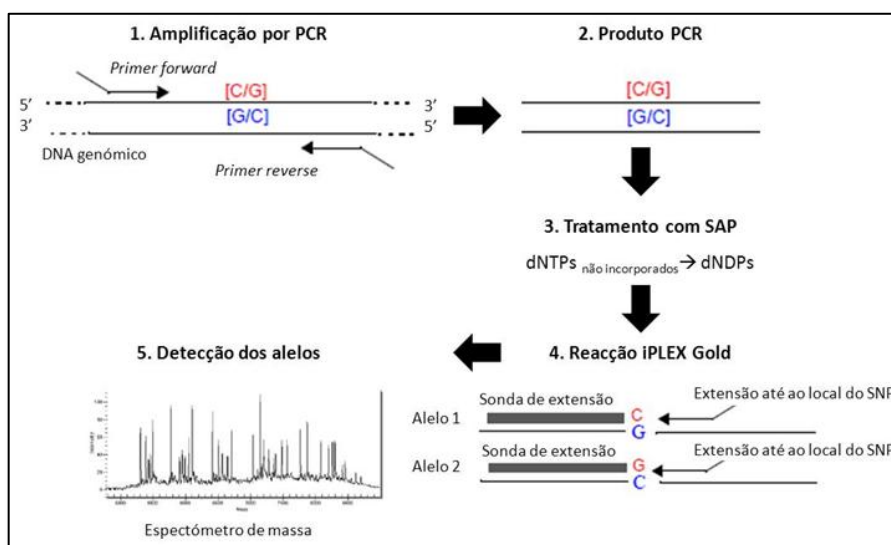


**Tabela 19** - Características e condições das enzimas utilizadas na genotipagem dos 4 SNPs por RFLP

SNP	Polimorfismo	Enzima	Local de restrição	Mistura de reacção (X1 reacção - µL)		Incubação
rs651539	CAAT[T/C]GGCC	TasI (Fermentas)	5' ... <sup>I</sup> AATT ... 3' 3' ... TTAA <sub>I</sub> ... 5'	10X Buffer B (Fermentas) Enzima Água Produto de PCR	1.0 0.1 3.9 8.0	65°C <i>overnight</i>
rs2975788	GAAG[G/A]TCCA	MboI (NEB)	5' ... <sup>I</sup> GATC ... 3' 3' ... CTAG <sub>I</sub> ... 5'	10X Buffer 4 (NEB) Enzima Água Produto de PCR	1.0 0.1 3.9 8.0	37°C <i>overnight</i>
rs17235018	ACAT[G/C]ATTC	Hinfi (Fermentas)	5' ... G <sup>I</sup> ANTC ... 3' 3' ... CTNA <sub>I</sub> G ... 5'	10X Buffer R (Fermentas) Enzima Água Produto de PCR	0.3 0.1 2.6 8.0	37°C <i>overnight</i>
rs4864056	TCTT[T/C]CGGG	Hpy188III (NEB)	5' ... TC <sup>I</sup> NNGA ... 3' ... 3' ... AGNN <sub>I</sub> CT ... 5'	10X Buffer 4 (NEB) Enzima BSA Água Produto de PCR	1.0 0.1 0.1 3.8 8.0	37°C <i>overnight</i>

### 3.3.4.2. Sequenom iPLEX Gold

«O Sequenom iPLEX Gold é um método universal para a detecção de inserções, deleções, substituições e outros polimorfismos, como SNPs, em DNA amplificado. Esta técnica é particularmente interessante pelo facto de permitir analisar diversos SNPs simultaneamente num mesmo ensaio. Para tal, é necessário desenhar *primers* muito específicos para cada fragmento de interesse e garantir que, nem a sequência dos *primers* se sobrepõem, nem as massas moleculares dos fragmentos. A utilização do *software* SEQUENOM's Assay Design facilita esta tarefa. No caso deste trabalho em particular foram feitas duas análises *multiplex*: uma continha 25 SNPs e a outra 20 SNPs. Os passos da técnica encontram-se ilustrados na Figura 10.

**Figura 10** – Representação esquemática das fases da técnica Sequenom iPLEX Gold

A primeira fase da técnica é um PCR convencional, no qual os fragmentos com os SNPs de interesse são amplificados. Um *robot* MassARRAY Liquid Handler (Sequenom, San Diego, California) distribui a mistura de reacção (Tabela 20) pelas placas (de 384 poços, no caso deste trabalho) previamente preparadas com o DNA seco (10ng/reacção) e de seguida, num termociclador convencional, sujeitam-se as amostras a um programa específico de amplificação (Tabela 21). Os *primers* utilizados nesta fase encontram-se ilustrados igualmente na Tabela suplementar 2.

**Tabela 20** – Quantidades e concentrações dos reagentes no volume final de reacção de PCR de Sequenom iPLEX Gold

Reagentes	X1 rx* ( $\mu$ l)	[final] em 5 $\mu$ l
10X PCR Buffer com 20mM MgCl <sub>2</sub>	0.5	1X (2.0mM MgCl <sub>2</sub> )
MgCl <sub>2</sub> (25mM)	0.4	2.0mM
dNTPs <sup>†</sup>	0.1	500 $\mu$ M
<i>Primers</i> ~	1.0	100nM
DNA polimerase (5U/ $\mu$ l)	0.2	1U/rx
Água ultrapura	1.8	n.a.
<b>Total</b>	<b>4.0</b>	

rx – reacção; \* a mistura de reacção deve ser preparada ligeiramente em excesso por forma a compensar possíveis erros de pipetagem; <sup>†</sup> foi feita uma mistura de dNTPs na qual a concentração de cada um era 25mM; ~ foi feita uma mistura de *primers* na qual a concentração de cada um era 500nM; n.a. – não se aplica.

**Tabela 21** – Programa utilizado nas PCR dos SNPs a analisar por Sequenom iPLEX Gold

Etapas	Intervalo de tempo	Temperatura (°C)	Nº ciclos
Desnaturação inicial	2'	94°C	45
Desnaturação	30"	94°C	
<i>Annealing</i>	30"	56°C	
Extensão	1'	72°C	
Extensão final	1'	72°C	
Pausa	$\infty$	4°C	

Após a amplificação dos fragmentos, é necessário proceder a uma neutralização dos dNTPs que não foram incorporados nas cadeias de DNA sintetizadas, convertendo-os em difosfatos de desoxirribonucleótidos (dNDPs), a fim de os tornar disfuncionais em reacções futuras. Assim, aos produtos de PCR adiciona-se, recorrendo novamente a um *robot* MassARRAY Liquid Handler (Sequenom, San Diego, California), uma solução contendo a enzima SAP (*Shrimp Alkaline Phosphatase* – fosfatase alcalina de camarão) (Tabela 22) e incubam-se as placas (Tabela 23).

**Tabela 22** – Quantidades dos reagentes na mistura de reacção de neutralização de Sequenom iPLEX Gold

Reagentes	X1 reacção* (µl)
SAP Buffer (10X)	0.17
Enzima SAP (1.7U/µl)	0.30
Água ultrapura	1.53
<b>Total</b>	<b>2.0~</b>

rx – reacção; \* a mistura de reacção é preparada em excesso por forma a compensar possíveis erros de pipetagem; ~ os 2.0µl/rx são adicionados ao produto de PCR (5.0µl), perfazendo um volume total de reacção de 7.0µl

**Tabela 23** – Condições da incubação da fase de neutralização de Sequenom iPLEX Gold

Etapas	Intervalo de tempo	Temperatura (°C)
<b>Passo 1</b>	40'	37°C
<b>Passo 2</b>	5'	80°C
<b>Pausa</b>	∞	4°C

Neutralizados os dNTPs, procede-se à reacção de extensão iPLEX Gold. Uma nova mistura de reacção é preparada (Tabela 24) e distribuída pelas placas contendo o produto de PCR e neutralização.

Novamente as amostras são sujeitas a um procedimento semelhante ao PCR (Tabela 25), mas no qual ocorre um *annealing* de sondas de extensão (que complementam o fragmento amplificado até ao nucleótido imediatamente anterior ao *locus* do SNP em análise) (Tabela suplementar 3), em vez dos tradicionais *primers*. A etapa da extensão difere igualmente do PCR convencional, na medida em que apenas um nucleótido é incorporado. Este é um nucleótido de terminação que marca o final do fragmento e permite a sua detecção.

**Tabela 24** – Quantidades dos reagentes na mistura de reacção de extensão de Sequenom iPLEX Gold

Reagentes	X1 reacção* (µl)	[final] em 9µl
iPLEX Buffer Plus (10X)	0.2000	0.222X
iPLEX Termination mix	0.2000	1X
<i>Primer mix (5µM; 10µM; 15µM)~</i>	0.9400	0.52µM; 1.04µM; 1.57µM*
Enzima iPLEX	0.0410	1X
Água ultrapura	0.6190	n.a.
<b>Total</b>	<b>2.0000</b>	

\* a mistura de reacção deve ser preparada ligeiramente em excesso por forma a compensar possíveis erros de pipetagem; ~ mediante os *primers* sejam *low*, *medium* ou *high mass*, assim terão concentrações distintas (respectivamente) na *mix* de *primers* e na mistura de reacção; n.a. – não se aplica

**Tabela 25** – Programa utilizado na etapa de extensão de Sequenom iPLEX Gold

Etapas	Intervalo de tempo	Temperatura (°C)	Nº ciclos	
Desnaturação inicial	30"	94°C	5	45
Desnaturação	5"	94°C		
Annealing	5"	52°C		
Extensão	5"	80°C		
Extensão final	3'	72°C		
Pausa	∞	4°C		

Antes de proceder à detecção dos alelos por espectroscopia de massa, é necessário dessalinizar os produtos de extensão e transferi-los para um SpectroCHIP® array, recorrendo ao MassARRAY nanodispenser (Sequenom, San Diego, California). O SpectroCHIP® array é então analisado pelo MassARRAY analyzer (Sequenom, San Diego, California), um espectrómetro de massa MALDI-TOF (*Matrix-Assisted Laser Desorption/Ionization - Time-Of-Flight*) desenhado especificamente para análise genómica. Por fim, os resultados são analisados através do software TyperAnalyzer.

#### 3.3.4.3. TaqMan® SNP Pre-Designed Genotyping Assay

Os TaqMan® SNP Pre-Designed Genotyping Assay são ensaios pré-desenhados, otimizados para a genotipagem de SNPs. Existem ensaios disponíveis no mercado para a genotipagem de cerca de 1.8 milhões de *loci* humanos, cuja funcionalidade é testada em cerca de 20 amostras.

O princípio teórico do procedimento é em tudo semelhante ao do TaqMan® SNP Drug Metabolism Genotyping Assay descrito no capítulo 2.3.3. Apenas as condições da mistura de reacção podem variar um pouco.

As condições das misturas de reacção e do programa utilizado nestes TaqMan® SNP Pre-Designed Genotyping Assay encontram-se, respectivamente, descritas nas Tabela 26 e Tabela 27.

Os polimorfismos e fluoróforos acoplados a cada alelo encontram-se descritos na Tabela 28.

**Tabela 26** – Condições da mistura de reacção utilizada na execução dos ensaios TaqMan® SNP Pre-Designed Genotyping Assay

Reagentes	X1 reacção (µl)
2X TaqMan Universal PCR Master Mix, No AmpErase UNG	2.500
40X SNP Genotyping Assay	0.125
Água DNase-free	2.375
<b>Volume Total</b>	<b>5.000</b>

\* a mistura de reacção deve ser preparada ligeiramente em excesso por forma a compensar possíveis erros de pipetagem

**Tabela 27** – Programa utilizado na execução dos ensaios *TaqMan*® *SNP Pre-Designed Genotyping Assay* para a genotipagem do SNP rs1229984

Etapas	Temperatura (°C)	Tempo	Nº de ciclos
Activação da enzima AmpliTaq Gold – HOLD	95	10'	1
Desnaturação	92	15"	50
Annealing/Extensão	60	1'30"	50

**Tabela 28** – SNPs genotipados recorrendo à técnica *TaqMan*® *SNP Pre-Designed Genotyping Assay* com respectivos polimorfismos e fluoróforos acoplados a cada alelo

SNP	Polimorfismo	Alelos		Cadeia analisada na técnica
		VIC	6-FAM	
rs16887451	A/C	A	C	R
rs10830772	A/G	A	G	R
rs2044068	T/C	A	G	F

F – forward; R – reverse.

### Controlo de qualidade

No caso das técnicas Sequenom iPLEX Gold e *TaqMan*® *SNP Pre-Designed Genotyping Assay* o controlo de qualidade do ensaio foi efectuado através da utilização de indivíduos de genótipo conhecido (indivíduos HapMap) como controlos positivos; amostras duplicadas na mesma e entre as várias placas como controlos inter-ensaio; amostras de água, sem DNA como controlos negativos; e verificação do Equilíbrio de Hardy-Weinberg (HWE) para cada SNP.

No caso dos SNPs genotipados por RFLP, o controlo de qualidade foi feito inicialmente através da utilização de indivíduos de genótipo conhecido (indivíduos HapMap) – controlos positivos, os quais foram substituídos por amostras populacionais com genótipo igual assim que possível. Como controlos negativos utilizaram-se reacções em que o DNA foi substituído por água. Foi ainda verificado o HWE.

A verificação do HWE permite inferir acerca da estabilidade da distribuição das frequências alélicas e genótípicas do SNP na população em geral entre gerações. Sempre que os seus valores são  $>0.05$  existe evidência da referida estabilidade e, no contexto de genotipagem, uma certeza de que não ocorreu nenhum erro no processo. Contudo, uma vez que nesta fase não existia um grupo de controlos, um valor de HWE  $<0.05$  poderia não significar qualquer erro de genotipagem, mas sim uma estratificação da população analisada no sentido da existência de uma associação entre um determinado SNP e um fenótipo que se pretendesse estudar (Lunetta, 2008).

### 3.3.5. Análise estatística dos resultados

Após a genotipagem individual dos 52 SNPs seleccionados, os resultados obtidos foram analisados recorrendo ao software R (v2.14.1) (R Development Core Team, 2006). Os sucessivos passos da análise foram executados recorrendo aos *packs* MASS e SNPassoc do *software*, os quais permitiram a realização de regressões logísticas binárias, que permitem testar a associação entre um genótipo e um fenótipo.

Validação dos resultados do GWAS em *pooled samples* através da genotipagem individual de cada SNP nos indivíduos constituintes dos *pools*

Inicialmente procedeu-se à validação dos 52 SNPs seleccionados. Para tal, cada SNP foi genotipado individualmente em cada amostra utilizada nas *pooled samples* (N=187). Foi feita uma análise de associação através de uma regressão logística binária, os SNPs com *p-value* <0.05 revelaram estar associados à recuperação e foram considerados para análise na amostra combinada.

Análise de associação entre os SNPs validados e o padrão de recuperação observado após o AVC

Os SNPs validados foram analisados na amostra combinada. Foi então realizada uma análise da amostra total (N=414). Posteriormente, repetiu-se a análise dos SNPs, considerando co-variáveis (presença de hipertensão, afasia, paresia e alterações de consciência aquando do AVC e ocorrência de complicações durante o internamento) associadas ao padrão de recuperação após um AVC isquémico ( $P \leq 0.01$ ).

A determinação de factores de confundimento e sua inclusão na análise estatística pretende ajustar os resultados obtidos, na medida em que entra em linha de conta com o peso de outras co-variáveis previamente associadas à recuperação após um AVC. Desta forma, a análise da relação genótipo-fenótipo corrige para resultados significativos que se devam à presença de factores de confundimento e não a uma verdadeira associação entre o genótipo e o fenótipo.

Todas as análises de associação foram realizadas recorrendo ao software R (v2.14.1) (R Development Core Team, 2006) e tiveram por base o modelo genético log-aditivo (uma variante do modelo aditivo), pois este permite analisar a relação entre as variáveis sem considerar nenhum tipo de dominância/recessividade entre os alelos do SNP. Desta forma, cada genótipo é considerado de forma independente, sem que o heterozigótico seja analisado como parte integrante do grupo de homozigóticos para um dos alelos (o que acontece quando se consideram os modelos genéticos de dominância ou recessividade). Assim, cada cópia de cada alelo é avaliada independentemente (Lunetta, 2008).

### 3.4. RESULTADOS

Em todos os resultados obtidos, os controlos positivos e inter-ensaio mostraram-se consistentes e os negativos não apresentaram qualquer anomalia.

Aquando da análise, o valor do *p-value* associado ao Equilíbrio de Hardy-Weinberg (HWE) no total das amostras utilizadas revelou-se sempre >0.05, pelo que foi possível concluir que a genotipagem foi, de facto, bem-sucedida.

### 3.4.1. Validação dos resultados do GWAS em *pooled samples* por genotipagem individual

Uma vez que a alelotipagem em *pooled samples* e respectivas réplicas pode introduzir algum erro experimental no cálculo das frequências alélicas associadas a cada SNP de cada grupo de indivíduos, apenas se consideraram SNPs com um valor de coeficiente de variação (CV) inferior a 15%, na tentativa de excluir todos aqueles que apresentavam as maiores diferenças entre as réplicas da alelotipagem. Tendo posteriormente sido aplicados mais critérios de selecção (referidos em 3.3.3.), obteve-se um total de 52 SNPs a analisar – os resultados preliminares da alelotipagem indicavam-nos como os mais promissores para validação do GWAS. Assim, analisaram-se os resultados da genotipagem dos 52 SNPs nos 187 indivíduos utilizados na construção das *pooled samples* ( $mRS=0$  e  $mRS\geq 3$ ) e compararam-se estes resultados com os obtidos previamente do GWAS.

Da genotipagem individual dos 52 SNPs seleccionados nos 187 pacientes obtiveram-se os resultados apresentados na Tabela 29.

Sabendo que um valor de *p-value* inferior a 0.05 revela uma associação entre um SNP e um fenótipo, neste caso de boa ou má recuperação, a um nível de significância de 5%, observa-se que foi possível validar 24 SNPs ( $0.001 < P < 0.049$ ).

4 SNPs revelaram uma frequência alélica para o alelo mais comum de 100%, o que traduz a ausência de polimorfismo nesses *loci* na população estudada. Ao verificar as *Minor Allele Frequencies* (MAFs) destes SNPs, observou-se que 2 deles (rs2703313 e rs1144046) apresentavam uma  $MAF=0$  na população HapMap CEU (população caucasiana europeia e de descendência europeia no Utah (EUA), descrita no projecto HapMap como aquela com a qual as características da população portuguesa mais se assemelham). Contudo, as frequências alélicas eram distintas entre os grupos analisados no GWAS, o que indica a existência dos 2 alelos na população. Porém, estes resultados não foram validados. Os restantes 2 SNPs (rs1562646 e rs4864056), por outro lado, tinham MAFs de 0.21 e 0.18, respectivamente, na população HapMap. Ambos os SNPs apresentaram-se polimórficos nas amostras em *pool*, porém, estes resultados não foram validados.

Apesar da comparação das frequências alélicas obtidas na genotipagem com as previamente descritas para a população CEU ser uma boa referência do sucesso da técnica utilizada, nem sempre estas se replicam totalmente, porque a população portuguesa apresenta desvios de frequências alélicas relativamente à população CEU, essencialmente devido à marcada migração que se verificou ao longo da história do país.

**Tabela 29** – Resultados da validação da alelotipagem para cada SNP por genotipagem individual

SNP	Cr	Posição	Polimorfismo	Alelo <sup>*</sup>	Amostra da validação Frequências alélicas		P
					mRS=0	mRS $\geq 3$	
rs7514276	1	217031796	G/T	G	55.2	57.6	<b>0.016</b>
rs17048803	1	217130725	C/T	T	76.7	64.6	<b>0.012</b>
rs687417	1	217170628	A/C	A	59.8	50.5	0.082
rs651539	1	234904056	T/C	T	89.0	82.1	0.070
rs4660095	1	239706677	T/C	T	78.7	63.6	<b>0.002</b>

rs10514631	2	19767453	T/C	T	83.7	75.8	0.057
rs2699164	2	36335706	C/T	T	88.5	82.3	0.104
rs17024145	2	39768185	A/G	A	69.0	58.6	<b>0.042</b>
rs12711674	2	124670233	A/C	A	80.5	67.7	<b>0.002</b>
rs1562646	2	186956727	A/G	A	100.0	100.0	*
rs2975788	2	241220531	G/A	A	59.9	54.7	<b>0.008</b>
rs615329	3	4102195	A/G	A	75.9	64.1	<b>0.017</b>
rs17235018	3	7187512	G/C	G	70.7	64.1	0.178
rs10049174	3	55116477	C/T	T	70.1	56.1	<b>0.003</b>
rs11706616	3	175626076	T/C	T	84.5	75.0	<b>0.025</b>
rs4916566	3	198083321	A/G	A	79.9	67.0	<b>0.007</b>
rs1479249	4	11852542	C/T	T	79.3	72.7	0.160
rs11933940	4	37238765	G/A	A	63.2	54.5	0.090
rs4864056	4	132827716	T/C	C	100.0	100.0	*
rs16887451	5	17954332	A/C	A	91.0	83.0	0.050
rs11747001	5	132440198	A/G	A	85.6	76.8	<b>0.032</b>
rs2876137	6	8347264	A/G	A	82.2	73.0	<b>0.045</b>
rs9350278	6	20853694	T/A	T	86.2	73.5	<b>0.003</b>
rs4710950	6	20911078	A/T	T	64.9	54.0	<b>0.028</b>
rs10253978	7	33521197	A/G	A	95.4	91.4	0.107‡
rs2617091	8	4438281	T/G	T	69.0	57.6	<b>0.013</b>
rs9314584	8	5758935	C/G	C	81.0	74.2	0.108
rs17075289	8	5791863	G/T	G	93.7	93.4	0.921‡
rs2703313	8	5805958	A/C	A	100.0	100.0	*
rs13258732	8	104179751	T/G	T	87.2	81.5	0.155
rs13249775	8	109494368	C/T	C	70.7	58.1	<b>0.012</b>
rs7860786	9	94285294	C/T	C	54.0	59.6	0.276
rs10903686	10	2190219	G/A	G	54.0	59.1	0.315
rs1243411	10	9009808	G/T	G	61.5	74.7	<b>0.005</b>
rs11001208	10	76363007	T/C	T	92.4	90.4	0.837
rs10830772	11	91158549	A/G	G	56.6	63.9	0.182
rs11175781	12	64167286	T/G	T	69.5	52.0	<b>0.001</b>
rs7980667	12	101641189	A/G	A	85.1	76.0	<b>0.031</b>
rs11849729	14	45564779	C/T	C	60.9	52.0	<b>0.017</b>
rs11159233	14	76295162	C/A	C	54.6	67.2	<b>0.009</b>
rs2044068	14	87597185	T/C	A	85.1	79.6	0.101
rs16949386	15	93049222	A/G	A	98.9	92.9	<b>0.002‡</b>
rs4888652	16	76195008	G/A	A	67.2	59.5	0.131
rs7207469	17	532225	T/G	T	89.7	82.3	0.067
rs6506157	18	4030656	A/G	A	79.3	70.0	<b>0.049</b>
rs1144046	18	48282979	G/A	G	100.0	100.0	*
rs6039208	20	8536762	A/G	A	87.4	80.3	0.075



rs6140752	20	8770308	T/G	T	81.6	75.8	0.164
rs6086654	20	8786376	C/T	C	75.6	68.9	0.165
rs742758	20	47486019	A/C	A	90.8	83.8	0.110
rs5960190	X	56651979	T/C	T	69.0	54.5	<b>0.027</b>
rs17218233	X	125146041	A/G	A	81.6	77.8	0.474

\* o alelo considerado é o alelo ancestral; \*- SNPs monomórficos, o valor do *p*-value não se aplica; ‡- *p*-value calculado segundo o modelo genético de co-dominância

Por forma a inferir acerca da fiabilidade do método de alelotipagem utilizando *pooled samples*, procedeu-se ainda à comparação das diferenças entre as frequências alélicas dos SNPs validados nos dois grupos (mRS=0 e mRS≥3) quando estimadas a partir do GWAS ou a partir da genotipagem individual. A realização de um teste de Wilcoxon ( $P=0.149$ ) permitiu concluir que o GWAS subestima as diferenças entre as frequências alélicas dos grupos, mas as diferenças entre as frequências alélicas dos grupos calculadas após alelotipagem e genotipagem individual não são estatisticamente significativas (Figura 11).

A validação dos resultados do GWAS permite ainda inferir acerca da adequação do método de selecção de SNPs para análise aprofundada após o rastreio genómico e sua análise preliminar.

Analisando as percentagens de SNPs seleccionados através de cada método, verifica-se que 51.4% dos SNPs seleccionados com base na posição no *top 1000* foram validados, bem como 54.5% dos que foram escolhidos através de *clusters* de 3 ou mais SNPs separados entre si por menos de 100kb e 42.9% dos escolhidos por constituírem *clusters* de SNPs no mesmo gene. Desta forma, verifica-se que as 3 estratégias adoptadas para a selecção de SNPs a genotipar individualmente foram adequadas, ainda que o melhor critério seja, aparentemente, o dos *clusters* de SNPs separados por <100kb. Contudo, devido ao reduzido número de SNPs seleccionados segundo este critério, a afirmação anterior é discutível.

### 3.4.2. Análise de associação dos 24 SNPs validados na amostra combinada

Analisaram-se os resultados da genotipagem dos 24 SNPs validados no conjunto amostral completo (N=414).

Inicialmente não foi incluído na análise nenhum factor de confundimento, pois o primeiro propósito desta análise era

o de verificar se os resultados obtidos na validação dos SNPs e relativos a grupos com índices extremos de recuperação se mantinham consistentes quando se alargava a amostra e se incluíam índices de recuperação intermédios. O *cut-off* para o sucesso da recuperação foi definido ao nível do mRS entre 1 e 2, já que este é

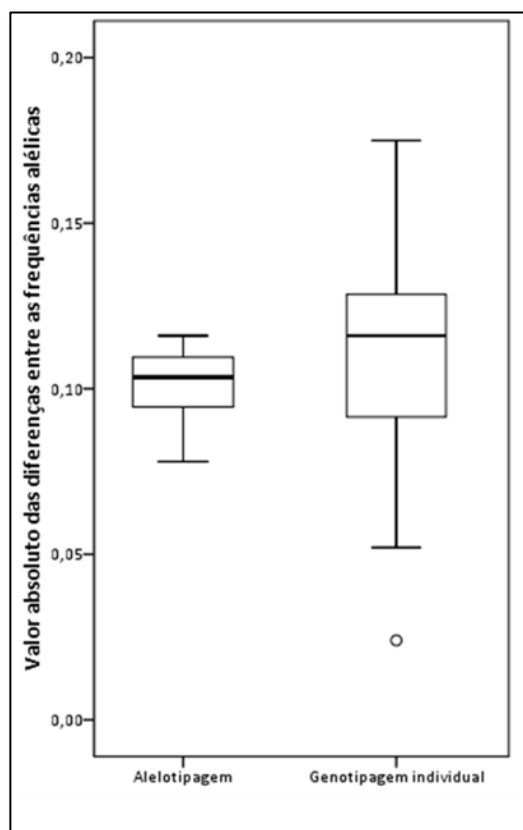


Figura 11 – Diferenças entre as frequências alélicas dos grupos calculadas após alelotipagem e genotipagem individual.

um valor clinicamente mais sensível. Posteriormente foi repetida a análise estatística, considerando agora os factores de confundimento que já se haviam revelado significativamente associados ao padrão de recuperação após um AVC isquémico (Tabela 14): presença de hipertensão, afasia, paresia e alterações de consciência aquando do AVC e ocorrência de complicações durante o internamento. Os resultados obtidos são apresentados na Tabela 30.

Tal como se pode observar precisamente na Tabela 30, 9 SNPs mantiveram o grau de significância quando se analisou a amostra combinada sem considerar factores de confundimento ( $0.01 < P < 0.041$ ).

O SNP rs2617091 foi o único genotipado de um *cluster* de SNPs localizados no gene *CSMD1* em que não foi possível inferir acerca de LD, dado o tamanho do gene. Dado o seu resultado significativo nesta fase e tendo em conta que a proteína codificada por este gene está envolvida em processos de supressão tumoral (Ma *et al.*, 2009), será interessante, futuramente, analisar os restantes SNPs do gene, apesar da sua posição no *top 1000* (posição 978).

Após inclusão na análise de factores de confundimento, o SNPs rs9350278 manteve um *p-value* associado significativo. Este SNP localiza-se no gene *CDKAL1* e encontra-se em LD com o SNP 9295481. Desta forma, ter-se-ão identificado 2 SNPs associados à recuperação após um AVC isquémico.

### 3.5. DISCUSSÃO

Contrariamente ao que se verifica em relação à susceptibilidade ao AVC, existem relativamente poucos estudos de associação acerca de factores genéticos que possam estar envolvidos na recuperação após um evento e a sua maioria tem sido desenvolvida testando genes candidatos e não numa abordagem genómica. Para contribuir para a compreensão das vias bioquímicas envolvidas na recuperação após um AVC isquémico, foi realizado um GWAS em *pooled samples* com o objectivo de identificar variantes genéticas associadas à recuperação dos pacientes. A metodologia adoptada (*pooled samples*) permite ir de encontro ao objectivo do trabalho, sendo custo-efectiva e abrangente, ou seja, não necessitando do estabelecimento prévio de uma hipótese, rastreando todo o genoma e potenciando a identificação de regiões de interesse que não seriam identificadas se testados apenas genes candidatos.

Numa fase inicial do estudo foram analisados SNPs com CV<10% entre réplicas, na tentativa de diminuir a probabilidade de encontrar falsos positivos. Nessa fase, a análise de associação ajustada aos parâmetros de severidade do AVC significativamente associados à recuperação após um AVC isquémico permitiu identificar 6 polimorfismos (SNPs) envolvidos nas vias de recuperação. 4 dos referidos SNPs (rs7664979, rs9293983, rs290916, rs1243659) localizam-se em regiões intergénicas, pelo que a associação com a recuperação sugere que na sua vizinhança (até 1Mb de distância) existam variantes funcionais igualmente associadas à recuperação. Os restantes 2 SNPs, rs10974334 e rs10273634 localizam-se em regiões intrónicas dos genes do cromossoma 9 *GLIS3* e *BBS9*, respectivamente.

**Tabela 30** – Resultados do teste de associação dos SNPs validados na amostra combinada (N=414)

SNP	Cr	Posição	Polimorfismo	Alelo*	HWE	Amostra combinada		P-value não ajustado	P-value ajustado	OR [95%CI]
						mRS≤1	mRS>1			
						Freq, Alélica (%)	Freq, Alélica (%)			
rs7514276	1	217031796	G/T	G	0.767	50.4	56.3	0.089	0.793	0.95[0.67-1.36]
rs17048803	1	217130725	C/T	T	0.347	72.7	66.6	0.063	0.915	0.98[0.67-1.43]
rs4660095	1	239706677	T/C	T	1.000	70.2	64.5	0.080	0.123	1.33[0.93-1.91]
rs17024145	2	39768185	A/G	A	0.914	66.3	62.4	0.250	0.175	1.29[0.89-1.87]
rs12711674	2	124670233	A/C	A	0.193	69.1	72.3	0.305	0.148	0.74[0.49-1.11]
rs2975788	2	241220531	G/A	A	0.090	56.1	50.8	0.061	0.117	1.33[0.93-1.89]
rs615329	3	4102195	A/G	A	0.903	78.7	67.7	<b>0.011</b>	0.237	1.27[0.85-1.89]
rs10049174	3	55116477	C/T	T	0.529	64.3	59.8	0.176	0.309	1.22[0.83-1.79]
rs11706616	3	175626076	T/C	T	0.764	82.4	75.5	<b>0.014</b>	0.480	1.17[0.75-1.83]
rs4916566	3	198083321	A/G	A	0.442	78.4	68.6	<b>0.002</b>	0.138	1.36[0.91-2.03]
rs11747001	5	132440198	A/G	A	0.747	84.1	77.7	<b>0.017</b>	0.066	1.54[0.97-2.43]
rs2876137	6	8347264	A/G	A	0.065	79.6	75.5	0.186	0.286	1.25[0.83-1.89]
rs9350278	6	20853694	T/A	T	0.531	84.6	75.3	<b>0.001</b>	<b>0.022</b>	1.66[1.07-2.59]
rs4710950	6	20911078	A/T	T	0.537	63.7	56.6	<b>0.034</b>	0.129	1.33[0.92-1.92]
rs2617091	8	4438281	T/G	T	0.397	67.8	57.7	<b>0.002</b>	0.195	1.28[0.88-1.86]
rs13249775	8	109494368	C/T	C	0.831	63.0	63.0	0.785	0.896	1.03[0.71-1.49]
rs1243411	10	9009808	G/T	G	0.911	63.3	73.1	<b>0.003</b>	0.067	0.70[0.47-1.03]
rs11175781	12	64167286	T/G	T	1.000	59.8	55.7	0.237	0.284	1.23[0.84-1.79]
rs7980667	12	101641189	A/G	A	0.096	82.6	80.9	0.538	0.232	1.31[0.84-2.04]
rs11849729	14	45564779	C/T	C	0.073	56.7	52.7	0.271	0.054	1.39[0.99-1.95]
rs11159233	14	76295162	C/A	C	0.266	56.1	61.4	0.132	0.095	0.75[0.53-1.05]
rs16949386	15	93049222	A/G	A	0.173	97.0	94.0	0.088	0.082	2.10[0.91-4.85]
rs6506157	18	4030656	A/G	A	0.395	86.9	76.2	0.409	0.574	1.13[0.74-1.71]
rs5960190	X	56651979	T/C	T	0.269	67.4	58.5	<b>0.041</b>	0.909	1.02[0.76-1.37]

Cr – cromossoma; HWE – Equilíbrio de Hardy-Weinberg; \* o alelo considerado é o alelo ancestral

Apesar dos resultados da análise inicial estabelecendo o *cut-off* para o CV ao nível dos 10%, considerou-se que seria interessante alargar o CV e repetir a análise. De facto, após inferir acerca de um fenótipo de recuperação mediado pelos genes *GLIS3* e *BBS9*, observou-se que um SNP de cada gene, rs7857436 e rs10486524, respectivamente, tinham sido excluídos desta primeira análise por apresentarem um CV ligeiramente superior a 10%, apesar de apresentarem diferenças entre as frequências alélicas nos grupos de boa e má recuperação semelhantes a outros SNPs analisados. No caso destes SNPs em particular, teria sido

muito interessante analisá-los, por forma a dar mais consistência à relação entre os genes em que se encontravam e o padrão de recuperação de um paciente após um AVC isquémico.

Assim, na segunda fase do estudo, concretizada no trabalho agora apresentado, aumentou-se o valor do *cut-off* para o CV para 15% e repetiu-se a análise. Este incremento poderá ter levado à inclusão de falsos positivos na análise (que poderão ter correspondido aos SNPs não validados na genotipagem individual), mas terá igualmente conduzido ao decréscimo do risco de exclusão de marcadores que pudessem estar associados ao fenótipo em estudo (falsos negativos).

A análise de associação não ajustada aos parâmetros de severidade do AVC significativamente associados à recuperação após um AVC isquémico permitiu identificar 24 SNPs. Porém, após o ajuste aos referidos parâmetros, apenas 1 polimorfismo (SNP) se revelou associado à recuperação. Este SNP, rs9350278 encontra-se em LD com o rs9295481 (pré-seleccionado para análise mas não genotipado precisamente por ser redundante, tal como explicado em 3.3.3), pelo que é possível inferir que também este último estará associado à recuperação após um AVC isquémico. Ambos os SNPs estão localizados no gene *CDKAL1* e, como tal, será interessante investigar se este gene será um potencial gene candidato para o fenótipo.

### CDKAL1

A proteína CDKAL1 (*CDK5 Regulatory Subunit Associated Protein 1-Like 1*) tem sido associada ao aumento do risco de diabetes tipo-2. De facto, esta proteína, homóloga da CDK5RAP1, inibe a acção da proteína *Cyclin-Dependent Kinase* (CDK5), responsável pela tradução de sinais de glucotoxicidade nas células  $\beta$  pancreáticas. Estudos em ratinhos *knock-out* para o gene *CDKAL1* têm permitido confirmar que, de facto, nestes animais os níveis de glucose no sangue encontram-se aumentados, a libertação de insulina ocorre de forma anormal e os ilhéus de Langerhans encontram-se hipertrofiados. (Mao *et al.*, 2012; McMurray *et al.*, 2012).

O efeito adverso da mutação do gene *CDKAL1* relacionado com a diabetes tipo-2 pode ainda ser agravado por uma dieta rica em gorduras (McMurray *et al.*, 2012). Estudos relativos à associação entre SNPs deste gene e o índice de massa corporal (IMC) têm sido realizados e têm mostrado que existirá de facto uma associação entre o genótipo e o fenótipo (Okada *et al.*, 2012).

Factores de risco (sejam genéticos, como é este caso em particular, ou ambientais) para a diabetes, poderão ter um papel importante na recuperação dos pacientes que sofrem AVCs isquémicos, pois esta patologia está associada a um pior prognóstico (Barrett *et al.*, 2009; Dupuis *et al.*, 2010; Kaarisalo *et al.*, 2005).

Também em indivíduos com diabetes diagnosticada, seria interessante investigar a hipótese dos SNPs do gene *CDKAL1* estarem associados a uma pior recuperação.

Para além da identificação de um possível novo gene candidato para a recuperação dos pacientes de AVC isquémico, uma análise mais abrangente do trabalho realizado permite ainda consolidar resultados previamente obtidos.

Tal como referido no subcapítulo 3.3.3, alguns SNPs não foram nesta fase do trabalho seleccionados para genotipagem individual porque já tinham sido genotipados previamente ou porque estavam em LD com SNPs que já tinham sido genotipados anteriormente. Por exemplo, os SNPs rs10273634 (associado à

recuperação) e rs10486524 (em *LD* com o anterior) do gene *BBS9* foram seleccionados nesta fase do trabalho, pelo que se considera que mais um marcador do gene *BBS9* está associado ao fenótipo em estudo. A consistência na selecção de SNPs deste gene segundo critérios distintos e a significância dos resultados reforça a associação deste gene com a recuperação após um AVC isquémico.

Em conclusão, o estudo no seu global fornece evidência de associação de três genes, *GLIS3*, *BBS9* e *CDKAL1* com a recuperação.

Não obstante dos resultados positivos acima apresentados e da identificação dos genes *GLIS3*, *BBS9* e *CDKAL1* como genes candidatos envolvidos na recuperação após um AVC isquémico, a replicação do estudo em populações independentes da analisada é necessária para validação das conclusões. Adicionalmente, será interessante proceder ao estudo alargado das diversas variantes genéticas de cada gene, nomeadamente ao estudo de SNPs funcionais, com o objectivo de compreender se estes condicionam alterações funcionais de vias e processos biológicos envolvidos na recuperação após um AVC isquémico.

Para além do interesse e importância do delineamento de novos estudos, também a replicação deste tipo de trabalhos é fundamental para colmatar falhas decorrentes da heterogeneidade das populações. De facto, a prevalência dos alelos de cada SNP entre populações é bastante variável e apenas uma amostra alargada permite obter resultados fiáveis, sem que se incorra num viés devido a um número insuficiente de indivíduos na amostra que impossibilite a visualização do efeito de um alelo num fenótipo.

Comparando as duas fases deste estudo, a primeira etapa (considerando o *cut-off* de 10% para o CV) apresentou uma taxa de validação de SNPs de 41% e a segunda (considerando o *cut-off* de 15% para o CV) de 50%. Porém, a primeira fase permitiu a associação de 6 SNPs (em 88 – 6.8%) com o fenótipo, enquanto a segunda fase apenas permitiu a associação de 1 SNP (em 48 – 2.1%) com o fenótipo. Contudo, sabe-se que não foram considerados na segunda fase do estudo alguns SNPs associados à recuperação por já terem sido analisados na primeira fase. Não tendo ocorrido a referida primeira fase, esses SNPs seriam seleccionados para genotipagem individual e obter-se-iam mais resultados positivos que contribuiriam para o aumento da taxa de validação e do número de SNPs identificados associados à recuperação para um *cut-off* de CV de 15%.

Dada a existência de resultados relevantes provenientes de ambas as etapas do estudo, seria ideal, em termos futuros, adaptar os critérios adoptados por forma a optimizar a selecção de SNPs e, dessa forma, a eficiência do estudo. Assim, sugere-se que, acima de tudo, se ajuste (entre 10% e 15%) o valor do *cut off* do CV mediante a metodologia de selecção dos SNPs.

Na fase de validação dos SNPs por genotipagem individual na amostra utilizada para a constituição das *pooled samples*, foi possível inferir acerca do sucesso de cada metodologia de selecção de SNPs. Porém, os resultados finais da análise de associação não se apresentaram coerentes com o sucesso de cada metodologia referido anteriormente. Após uma observação cuidada, as maiores diferenças entre as frequências alélicas nos grupos de *pooled samples* – *top 1000* – foi a estratégia através da qual foi possível

identificar mais variantes genéticas associadas à recuperação, com 3 SNPs do *top 1000* (considerando o *top 1000* delineado na primeira fase do trabalho (CV<10%)) a revelarem resultados estatisticamente significativos. Tendo em conta uma alteração do *cut-off* do CV, possivelmente os SNPs seleccionados segundo este critério com resultados significativos ao nível de 10% deixariam de se localizar no topo do *top 1000*, o qual passaria a ser ocupado por SNPs que, presumivelmente, representariam falsos positivos. Os SNPs realmente relevantes passariam então a ocupar uma posição na qual não seriam seleccionados, sendo necessário alargar bastante o número de SNPs a analisar. Assim, na tentativa de seleccionar SNPs com base nas diferenças entre as frequências alélicas nos grupos de *pooled samples*, o *cut off* do CV que aparenta ser mais adequado continua a ser o de 10%.

Com o propósito de estudar SNPs localizados em *cluster* num mesmo gene, a fim de identificar novos genes candidatos associados a um fenótipo, já seria então importante aumentar o *cut-off* do CV, uma vez que se encontraram neste trabalho resultados significativos até um CV≈12.5% (os valores do CV dos SNPs localizados em genes associados à recuperação assumem valores até 12.5% - Tabela 15). Simultaneamente ao aumento do referido *cut off*, também o estabelecimento de um *top 1000* teria que ser abolido, sob pena de exclusão de SNPs potencialmente interessantes. De facto, ao aumentar o *cut off* do CV para 15%, o *cluster* de SNPs localizados no gene *GLIS3*, que se concluiu ser um gene associado à recuperação, ficou excluído do *top 1000*.

Os melhores parâmetros a adoptar no estudo de SNPs seleccionados pela sua localização em *clusters* de SNPs separados por <100kb são impossíveis de inferir a partir da segunda fase deste estudo. Ao contrário do que é sugerido por Abraham e colaboradores (2008) e do que se verificou na fase de validação de SNPs em que esta estratégia parecia ser a mais efectiva, a análise de *clusters* de SNPs separados por <100kb não permitiu a identificação de qualquer associação entre um SNP e a recuperação após um AVC isquémico. Desta forma, sugere-se a revisão da importância e da pertinência deste método e a adopção de critérios mais restritivos na tentativa de identificação de *clusters* de SNPs separados por <100kb.

Tal como referido em 3.2.1, na primeira fase do trabalho foi ainda aplicado um outro critério de selecção de SNPs: o valor do *p-value* de um teste *t-student* realizado para verificar a significância das diferenças entre as frequências alélicas nos dois grupos. Os SNPs seleccionados na segunda fase do estudo de acordo com este critério coincidiam na íntegra com os que já haviam sido seleccionados e analisados segundo o mesmo critério na primeira etapa do trabalho. Ainda que só 2 SNPs tenham sido identificados como potencialmente envolvidos no padrão de recuperação após um AVC isquémico, mas uma vez que não se observaram quaisquer diferenças nos SNPs seleccionados através deste método entre as fases do estudo, sugere-se que se continue a adoptar esta estratégia de selecção de SNPs sem qualquer alteração do valor do CV.

Dado o panorama das metodologias utilizadas na selecção de SNPs a genotipar individualmente e principalmente devido ao número reduzido de SNPs que se revelaram de facto associados ao padrão de recuperação dos pacientes de AVC isquémico, continua a ser difícil indicar uma como a melhor. No entanto, é possível concluir acerca do melhor valor de *cut-off* do CV a aplicar futuramente em estudos que envolvam a análise de *pooled samples*. Partindo da premissa que é necessária e possível (em termos de custos) uma validação extensa dos resultados obtidos a partir de *pooled samples*, o ideal será adoptar um *cut-off* para o

CV ao nível dos 15%. Mantendo os critérios de selecção de SNPs e alargando o *top 1000* adoptado neste estudo, um *cut-off* de 15% incluirá os resultados mais interessantes observados ao nível de 10% e permitirá a identificação de novas variantes genéticas que ao nível de 10% não seriam estudadas.

Observando e analisando o primeiro passo deste estudo, a estratégia de GWAS em *pooled samples* e comparativamente a estudos anteriores que a utilizaram igualmente, verificou-se que neste trabalho a eficiência obtida (dada pela percentagem de SNPs validados) foi a mais elevada (cerca de 41% na primeira fase do estudo e 50% na segunda), provavelmente devido ao *design* experimental traçado, que incluiu réplicas de formação do *pool* e réplicas da técnica (Pearson *et al.*, 2007; Macgregor, 2007). De facto, Macgregor (2007) e Pearson e colaboradores (2007) sugerem que as réplicas são fundamentais para controlar variações associadas aos resultados. Estudos semelhantes levados a cabo por Butcher, Meaburn e respectivos colaboradores (2008), nos quais não foi elaborada qualquer réplica ao longo do trabalho, revelaram taxas de eficiência baixas (16% e 12%, respectivamente), enquanto outros estudos nos quais já foram realizadas réplicas da técnica revelaram taxas de eficiência superiores, de 17% e 31% (Shifman *et al.*, 2008a; Shifman *et al.*, 2008b).

Tal como é possível concluir a partir deste estudo, o sucesso da recuperação após um AVC isquémico é um processo complexo influenciado por uma importante componente genética. Contudo, é importante não esquecer que também existem factores ambientais a condicionar a referida recuperação. A identificação rápida do AVC e o acesso aos cuidados de saúde necessários (durante e após o evento) são determinantes na evolução do quadro clínico do paciente. De facto, apesar das variantes genéticas envolvidas, um bom aconselhamento médico, um acompanhamento emocional devido e uma reabilitação física adequada podem não só acelerar a recuperação como reduzir o risco de reincidência de um AVC (Ringleb *et al.*, 2011).

## **CAPÍTULO 4. CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS**



Concluído este estudo, foi possível observar o envolvimento de um polimorfismo do gene *ADH1B*, implicado no metabolismo do álcool, na susceptibilidade ao AVC, o que não é surpreendente, uma vez que o álcool é um importante factor de risco para o AVC. Contudo, alguns aspectos da análise revelaram-se inconclusivos devido a limitações amostrais. Desta forma, sugere-se a repetição desta análise, partindo de novos pressupostos, nomeadamente de uma amostra populacional melhor caracterizada nomeadamente em relação aos hábitos de consumo de álcool e de maior dimensão. O trabalho em curso a ser desenvolvido pelo grupo *Genetic Epidemiology*, Londres, colmatará a questão da dimensão da amostra. Estando a ser desenvolvido em maior escala, permitirá tirar conclusões mais fiáveis acerca do papel do SNP rs1229984 sobre os fenótipos de doença cardiovascular como o AVC.

Adicionalmente, seria interessante proceder a estudos fisiológicos, por forma a entender exactamente quais as implicações do consumo de álcool sobre diversas vias fisiológicas potencialmente envolvidas no AVC. Desta forma pretender-se-ia compreender quais as doses de álcool consideradas preventivas e poder-se-ia proceder a uma educação das populações mais específica, no sentido de alertar para os benefícios de um consumo regrado e para os malefícios do consumo excessivo de álcool, o qual é um problema grave a nível mundial.

Relativamente à recuperação após um AVC isquémico, foi possível confirmar que existe de facto uma importante componente genética envolvida neste processo. Apesar da importância dos factores ambientais como, por exemplo, a rapidez na obtenção de cuidados médicos e os planos de reabilitação física no sucesso da recuperação, também o conhecimento dos factores genéticos envolvidos permitirá melhorar a abordagem ao paciente aquando de um AVC. De facto, através do conhecimento das vias fisiológicas implicadas na recuperação será possível melhorar o prognóstico, bem como ajustar a terapêutica e a prevenção secundária. Assim, sugere-se a realização de estudos futuros, nomeadamente estudos funcionais relativos aos genes identificados como genes candidatos envolvidos na recuperação, a fim de se atingir o objectivo de saber como melhorar a qualidade de vida de um paciente que sofreu um acidente vascular cerebral.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G (2008) A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* 1:44
- Adibhatla RM, Hatcher JF (2008) Tissue plasminogen activator (tPA) and matrix metalloproteinases in the pathogenesis of stroke: therapeutic strategies. *CNS Neurol Disord Drug Targets* 7:243-253
- Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG (2009) Classification of Stroke Subtypes. *Cerebrovasc Dis* 27:493–501
- Andlin-Sobocki P, Jönsson B, Wittchen HU, Olesen J (2005) Costs of Disorders of the Brain in Europe. *Eur J Neurol* 12(Suppl.1)
- Applied Biosystems (2010) TaqMan® SNP Genotyping Assays – Protocol. Applied Biosystems, Foster City
- Atochin DN, Yuzawa I, Li Q, Rauwerdink KM, Malhotra R, Chang J, Brouckaert P, Ayata C, Moskowitz MA, Bloch KD, Huang PL, Buys ES (2010) Soluble guanylate cyclase  $\alpha 1\beta 1$  limits stroke size and attenuates neurological injury. *Stroke* 41:1815-1819
- Barrett JC, Clayton D, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS (2009) Genome-wide association study and meta-analysis finds over 40 loci affect risk of type 1 diabetes. *Nat Genet* 41(6):703–707
- Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ (2008) A genome-wide association study implicates diacylglycerol kinase  $\epsilon$  (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* 13(2):197-207
- Brass LM (1992) Stroke. In: Zaret BL, Berliner RW, Moser M, Cohen LS (1992) Yale University School of Medicine - Heart Book, 1<sup>st</sup> edition. Hearst Books, New York, pp 215-233
- Butcher LM, Davis OS, Craig IW, Plomin R (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav* 7:435-446
- Carr LG, Foroud T, Stewart T, Castelluccio P, Edenberg HJ, Li TK (2002) Influence of ADH1B Polymorphism on Alcohol Use and its Subjective Effects in a Jewish Population. *Am J Med Genet* 112:138–143
- Chang HW, Yang CH, Chang PL, Cheng YH, Chuang LY (2006) SNP-RFLPping: restriction enzyme mining for SNPs in genomes. *BMC Genomics* 7:30
- Chen CC, Lu RB, Chen YC, Wang MF, Chang YC, Li TK, Yin SJ (1999) Interaction between the Functional Polymorphisms of the Alcohol-Metabolism Genes in Protection against Alcoholism. *Am J Hum Genet* 65:795–807
- Cheung B, Holmes RS, Easteal S, Beacham IR (1999) Evolution of Class I Alcohol Dehydrogenase Genes in Catarrhine Primates: Gene Conversion, Substitution Rates, and Gene Regulation. *Mol Biol Evol* 16(1):23–36
- Consortium IH (2003) The International HapMap Project. *Nature* 426:789-796
- Correia M, Silva MR, Matos I, Magalhães R, Lopes JC, Ferro JM, Silva MC (2004) Prospective Community-Based Study of Stroke in Northern Portugal: Incidence and Case Fatality in Rural and Urban Populations. *Stroke* 35:2048-2053
- Davenport R, Dennis M (2000) Neurological emergencies: acute stroke. *J Neurol Neurosurg Psychiatry* 68:277–288
- Davidson College Biology Department (DCBD) (2010) Ozone and Our Lungs. Davidson College Biology Department. [http://www.bio.davidson.edu/people/kabernd/BerndCV/Lab/Online\\_website2/ethanol.html](http://www.bio.davidson.edu/people/kabernd/BerndCV/Lab/Online_website2/ethanol.html) (accessed 20 October 2012)
- Deb P, Sharma S, Hassan KM (2010) Pathophysiologic mechanisms of acute ischemic stroke: An overview with emphasis on therapeutic significance beyond thrombolysis. *Pathophysiology* 17:197–218
- DiCastelnuovo A, Costanzo S, Bagnardi V, Donati MB, Iacoviello L, Gaetano G (2006) Alcohol Dosing and Total Mortality in Men and Women: An Updated Meta-analysis of 34 Prospective Studies. *Arch Intern Med* 166:2437-2445

- Diergaarde B, Brand R, Lamb J, Cheong SY, Stello K, Barmada MM, Feingold E, Whitcomb DC (2010) Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. *Pancreatology* 10:194-200
- Domingues-Montanari S, Mendioroz M, Rio-Espinola A, Fernández-Cadenas I, Montaner J (2008) Genetics of stroke: a review of recente advances. *Expert Ver Mol Diagn* 8(4):495-513
- Drogan D, Sheldrick AJ, Schütze M, Knüppel S, Andersohn F, Giuseppe R, Herrmann B, Willich SN, Garbe E, Bergmann MM, Boeing H, Weikert C (2012) Alcohol Consumption, Genetic Variants in Alcohol Deydrogenases, and Risk of Cardiovascular Diseases: A Prospective Study and Meta-Analysis. *PLoS ONE* 7(2). doi:10.1371/journal.pone.0032176
- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparsø T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proença C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccascaccia RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllenstein U, Hadjadj S, et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42:105-116
- Edenberg HJ (2007) The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 30(1):5-13
- Ekman M (2004) Economic evidence in stroke: a review. *Eur J Health Econom Suppl* 1:74-83
- Gilhus NE, Barnes MR, Brainin M (2011) *European Handbook of Neurological Management – volume 1*. Blackwell Publishing, Oxford
- Gulcher JR, Gretarsdottir S, Helgadóttir A, Stefansson K (2005) Genes contributing to risk for common forms of stroke. *Trends Mol Med* 11(5):217-224
- Guo JM, Liu AJ, Su DF (2010) Genetics of stroke. *Acta Pharmacol Sin* 31: 1055–1064
- Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Anderson CS (2002) Long-Term Disability After First-Ever Stroke and Related Prognostic Factors in the Perth Community Stroke Study, 1989 -1990. *Stroke* 33:1034-1040
- Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C (2004) Ten-Year Risk of First Recurrent Stroke and Disability After First-Ever Stroke in the Perth Community Stroke Study. *Stroke* 35:731-735
- Humphries SE, Morgan L (2004) Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. *Lancet Neurol* 3:227–36
- Husemoen LLN, Fenger M, Friedrich N, Tolstrup JS, Fredriksen SB, Linneberg A (2008) The association of ADH and ALDH gene variants with alcohol drinking habits and Cardiovascular Disease risk factors. *Alcohol Clin Exp Res* 32(11):1984–1991
- IBM Corp (2011) *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, New York
- Imhof A, Woodward M, Doering A, Helbecque N, Loewel H, Amouyel P, Lowe GDO, Koenig W (2004) Overall alcohol intake, beer, wine, and systemic markers of inflammation in western Europe: results from three MONICA samples (Augsburg, Glasgow, Lille). *Eur Heart J* 25:2092–2100
- Ingall T (2004) Stroke—Incidence, Mortality, Morbidity and Risk. *J Insur Med* 36:143–152
- Jelski W, Szmítowski M (2008) Alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) in the cancer diseases. *Clin Chim Acta* 395:1–5
- Kaarisalo MM, Rähä I, Sivenius J, Immonen-Rähä P, Lehtonen A, Sarti C, Mähönen M, Torppa J, Tuomilehto J, Salomaa V (2005) Diabetes worsens the outcome of acute ischemic stroke. *Diabetes Res Clin Pract* 69:293–298

- Kasner SE (2006) Clinical interpretation and use of stroke scales. *Lancet Neurol* 5:603–12
- Kirshner HS (2009) Differentiating ischemic stroke subtypes: Risk factors and secondary prevention. *J Neurol Sci* 279:1–8
- Kloner RA, Rezkalla SH (2007) To Drink or Not to Drink? That Is the Question. *Circulation* 116:1306-1317
- Krawetz S, Misener S (2000) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa
- Lanktree MB, Dichgans M, Hegele RA (2010) Advances in Genomic Analysis of Stroke: What Have We Learned and Where Are We Headed?. *Stroke* 41:825-832
- Lisabeth LD, Smith MA, Brown DL, Uchino K, Morgenstern LB (2005) Family history and stroke outcome in a bi-ethnic, population-based stroke surveillance study. *BMC Neurol* 5:20
- Lloyd-Jones D, Adams R, Carnethon M, Simone G, Ferguson TB, Flegal K, Ford E, Furie K, Go A, Greenlund K, Haase N, Hailpern S, Ho M, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott M, Meigs J, Mozaffarian D, Nichol G, O'Donnell C, Roger V, Rosamond W, Sacco R, Sorlie P, Stafford R, Steinberger J, Thom T, Wasserthiel-Smoller S, Wong N, Wylie-Rosett J, Hong Y (2009) Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 119:e21-e181
- Lunetta KL (2008) Genetic Association Studies. *Circulation* 118:96-101
- Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, Cohen MA, Wang Y, Samanta M, Kumar MS, Aziz MU, Naylor TL, Weber BL, Fakharzadeh SS, Weinstein GS, Vachani A, Feldman MD, Brose MS (2009) Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. *Cancer Biol Ther* 8(10):907-916
- Macgregor S (2007) Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *Eur J Human Genet* 15:501–504
- Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PAF, Richter MM, Montgomery GW, Martin NG, Heath AC, Whitfield JB (2009) Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum Mol Gen* 18:580–593
- Maguire J, Thakkestian A, Levi C, Lincz L, Bisset L, Sturm J, Scott R, Whyte S, Attia J (2011) Impact of COX-2 rs5275 and rs20417 and GP1IIa rs5918 Polymorphisms on 90-Day Ischemic Stroke Functional Outcome: A Novel Finding. *J Stroke Cerebrovasc Dis* 20(2):134-144
- Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2010) Variants of the Matrix Metalloproteinase-2 but not the Matrix Metalloproteinase-9 genes significantly influence functional outcome after stroke. *BMC Med Genet* 11:40
- Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Variants in the inflammatory IL6 and MPO genes modulate stroke susceptibility through main effects and gene-gene interactions. *J Cereb Blood Flow Metab* 31(8):1751-1759
- Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2012) Variants within the nitric oxide synthase 1 gene are associated with stroke susceptibility. *Atherosclerosis* 220(2):443-448
- Mao H, Li Q, Gao S (2012) Meta-Analysis of the Relationship between Common Type 2 Diabetes Risk Gene Variants with Gestational Diabetes Mellitus. *PLoS ONE*. doi:10.1371/journal.pone.0045882
- Matarin M, Singleton A, Hardy J, Meschia J (2010) The genetics of ischaemic stroke. *J Intern Med* 267:139-155
- McMurray F, Moir L, Cox RD (2012) From Mice to Humans. *Curr Diab Rep*. doi: 10.1007/s11892-012-0323-2
- Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R (2008) Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. *Mol Psychiatry* 13:729–740
- National Center for Biotechnology Information, NCBI (2012) CDKAL1 CDK5 regulatory subunit associated protein 1-like 1 [Homo sapiens]. National Center for Biotechnology Information. <http://www.ncbi.nlm.nih.gov/gene/54901> (accessed 20 October 2012)

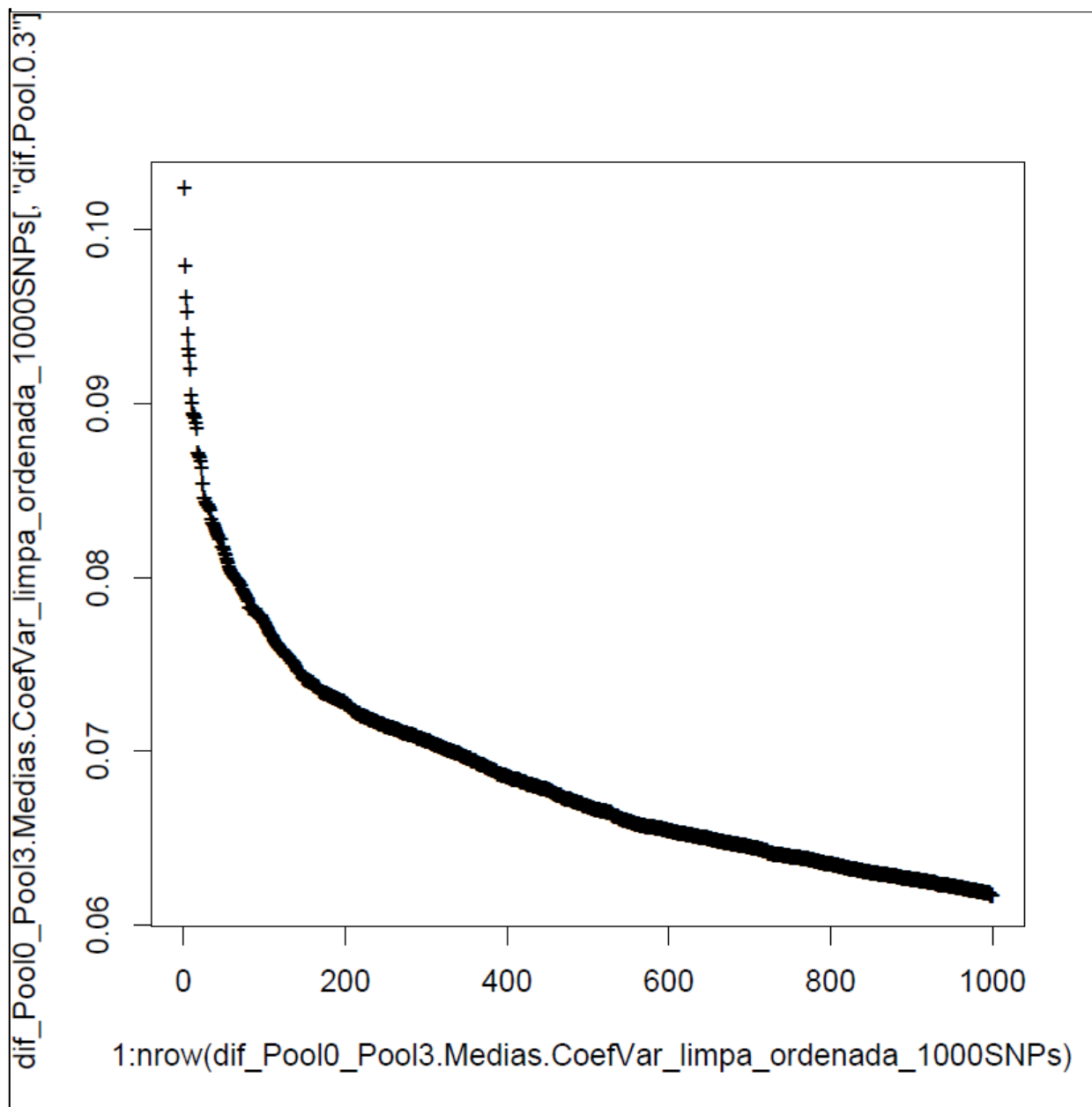
- National Institute of Neurological Disorders and Stroke (NINDS) (2004) Stroke: Hope Through Research. National Institute of Neurological Disorders and Stroke. [http://www.ninds.nih.gov/disorders/stroke/detail\\_stroke.htm](http://www.ninds.nih.gov/disorders/stroke/detail_stroke.htm) (accessed 20 October 2012)
- Okada Y, Kubo M, Ohmiya H, Takahashi A, Kumasaka N, Hosono N, Maeda S, Wen W, Dorajoo R, Go MJ, Zheng W, Kato N, Wu JY, Lu Q, Tsunoda T, Yamamoto K, Nakamura Y, Kamatani N, Tanaka T (2012) Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat Genet* 44(3):302-306
- Ovbiagele B, Saver JL (2010) Day-90 Acute Ischemic Stroke Outcomes Can Be Derived from Early Functional Activity Level. *Cerebrovasc Dis* 29:50–56
- Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, Webster JA, Beach T, Sando SB, Aasly JO, Heun R, Jessen F, Kölsch H, Tsolaki M, Daniilidou M, Reiman EM, Papassotiropoulos A, Hutton ML, Stephan DA, Craig DW (2007) Identification of the Genetic Basis for Complex Disorders by Use of Pooling-Based Genomewide Single-Nucleotide–Polymorphism Association Studies. *Am J Hum Genet* 80(1):126–139
- Peddareddygari LR, Dutra AV, Levenstien MA, Sen S, Grewal RP (2009) An analysis of methylenetetrahydrofolate reductase and glutathione S-transferase omega-1 genes as modifiers of the cerebral response to ischemia. *BMC Neurol* 9:37
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81
- R Development Core Team (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org> (accessed 20 October 2012)
- Rimm EB, Moats C (2007) Alcohol and Coronary Heart Disease: Drinking Patterns and Mediators of Effect. *Ann Epidemiol* 17:S3–S7
- Ringleb PA, Bousser M-G, Ford G, Bath P, Brainin M, Caso V, Cervera Á, Chamorro AL, Cordonnier C, Csiba L, Davalos A, Diener H-C, Ferro J, Hacke W, Hennerici M, Kaste M, Langhorne P, Lees K, Leys D, Lodder J, Markus HS, Mas J-L, Mattle HP, Muir K, Norrving B, Obach V, Paolucci S, Ringelstein EB, Schellinger PD, Sivenius J, Skvortsova V, Stibrant Sunnerhagen K, Thomassen L, Toni D, Von Kummer R, Wahlgren NG, Walker MF, Wardlaw J (2011) Ischaemic stroke and transient ischaemic attack. In: Gilhus NE, Barnes MR, Brainin M (2011) *European Handbook of Neurological Management – volume 1*. 2<sup>nd</sup> edition. Blackwell Publishing, Oxford, pp 101-158
- Roche (2006) PCR Applications Manual. Roche, Germany
- Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, pp 365-386
- Saenger AK, Christenson RH (2010) Stroke Biomarkers: Progress and Challenges for Diagnosis, Prognosis, Differentiation, and Treatment. *Clin Chem* 56(Suppl.1):21–33
- Senée V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, Charon C, Nicolino M, Boileau P, Cavener DR, Bougnères P, Taha D, Julier C (2006) Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet* 38:682-687
- Sequenom (2009) iPLEX Gold – Application Guide. Sequenom, San Diego
- Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hetttema JM, An SS, Neale MC, Oord EICG, Kendler KS, Chen X, Boomsma DI, Middeldorp CM, Hottenga JJ, Slagboom PE, Flint J (2008a) A whole genome association study of neuroticism using DNA pooling. *Mol Psychiatry* 13:302–312
- Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR1, Sun C, Flint J, Darvasi A (2008b) Genome-Wide Association Identifies a Common Variant in the Reelin Gene That Increases the Risk of Schizophrenia Only in Women. *PLoS Genet*. doi:10.1371/journal.pgen.0040028
- Silverman IE, Rymer MM, Ruilope LM (2011) *Visual Guide for Clinicians – Understanding Stroke*. Clinical Publishing, Oxford

- Truelsen T, Ekmanb M, Boysena G (2005) Cost of stroke in Europe. *Eur J Neurol* 12(Suppl.1):78–84
- Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 6(2):109–118
- Warlow C, Gijn JV, Dennis M, Wardlaw J, Bamford J, Hankey G, Sandercock P, Rinkel G, Langhorne P, Sudlow C, Rothwell P (2007) *Stroke – Practical Management*. Blackwell Publishing, Oxford.
- Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P (2003) Stroke. *Lancet* 362:1211–24
- Wilson JTL, Hareendran A, Grant M, Baird T, Schulz UGR, Muir KW, Bone I (2002) Improving the Assessment of Outcomes in Stroke: Use of a Structured Interview to Assign Grades on the Modified Rankin Scale. *Stroke* 33:2243–2246
- World Health Organization (WHO). (2011) Alcohol. World Health Organization. <http://www.who.int/mediacentre/factsheets/fs349/en/> (accessed 20 October 2012)
- Xi G, Keep RF, Hoff JT (2006) Mechanisms of brain injury after intracerebral haemorrhage. *Lancet Neurol* 5:53–63
- Yukawa Y, Muto M, Hori K, Nagayoshi H, Yokoyama A, Chiba T, Matsuda T (2012) Combination of ADH1B\*2/ALDH2\*2 polymorphisms alters acetaldehyde-derived DNA damage in the blood of Japanese alcoholics. *Cancer Sci* 1–5
- Zaret BL, Berliner RW, Moser M, Cohen LS (1992) *Yale University School of Medicine - Heart Book*. Hearst Books, New York

## **ANEXOS**

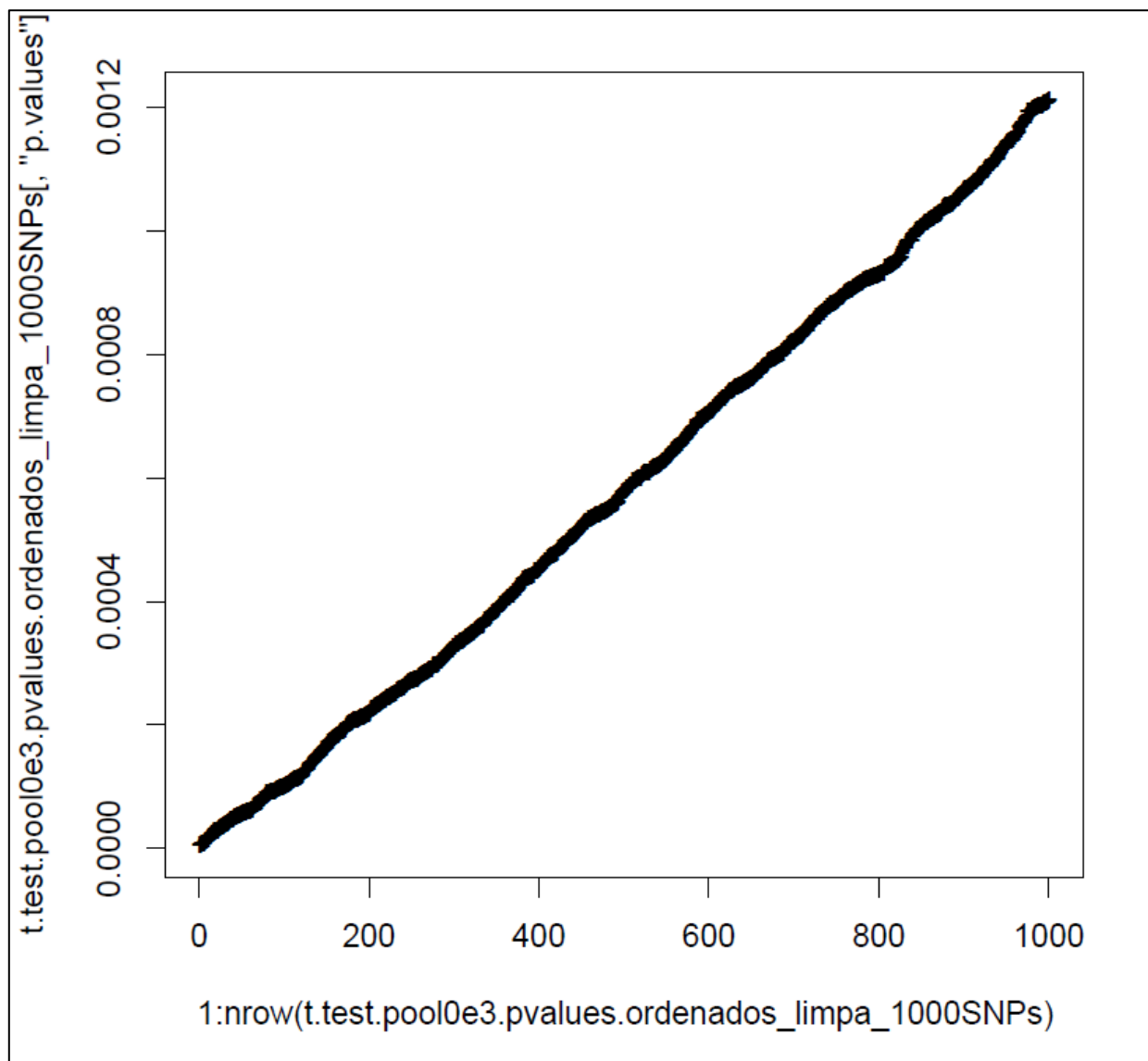
## FIGURAS SUPLEMENTARES

**Figura Suplementar 1** - Gráfico da distribuição do *top 1000* ( $CV < 10\%$ ). SNPs com maiores diferenças entre as frequências alélicas dos grupos ( $mRS=0$  e  $mRS \geq 3$ ) calculadas após a alelotipagem dos 262 264 analisados no GWAS.

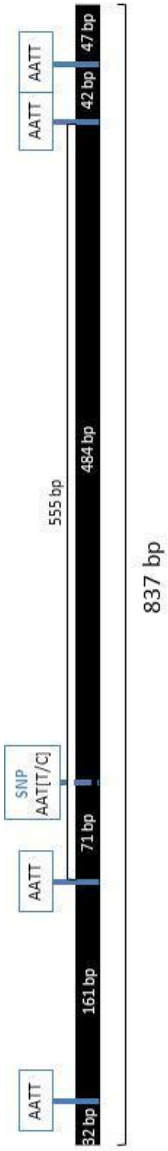
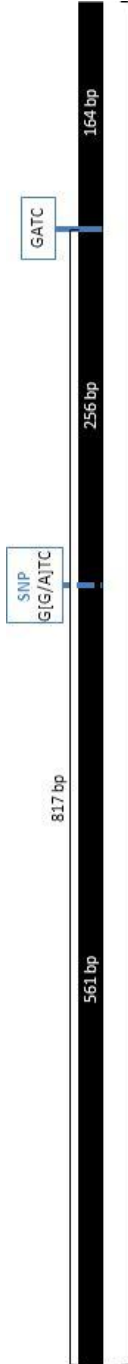
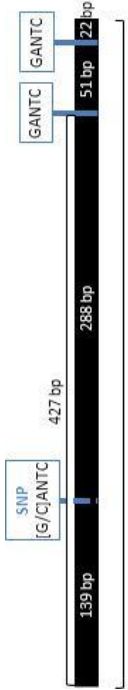
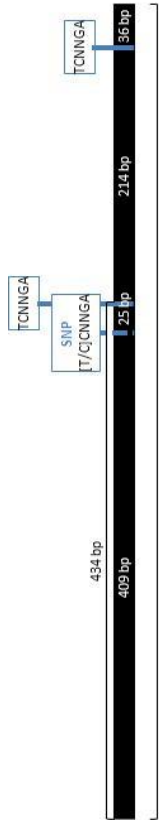




**Figura Suplementar 2** – Gráfico da distribuição dos 1000 SNPs com *p-values* mais baixos, após a realização de um teste *t* de Student para verificar a significância das diferenças entre as frequências alélicas nos dois grupos (boa recuperação – mRS=0 e má recuperação – mRS≥3)



**Figura Suplementar 3** – Representação esquemática dos locais de corte de enzimas de restrição utilizadas nas sequências amplificadas por PCR e padrão de fragmentos associados a cada genótipo

SNP	Esquema da restrição	Fragmentos obtidos
rs651539		Homozigotia alelo T: 32+161+ <b>71</b> +484+42+47 Homozigotia alelo C: 32+161+ <b>555</b> +42+47 Heterozigotia: 32+161+ <b>71</b> +484+ <b>555</b> +42+47
rs2975788		Homozigotia alelo G: <b>817</b> +164 Homozigotia alelo A: <b>561</b> + <b>256</b> +164 Heterozigotia: <b>817</b> + <b>561</b> + <b>256</b> +164
rs17235018		Homozigotia alelo G: <b>139</b> + <b>288</b> +51+22 Homozigotia alelo C: <b>427</b> +51+22 Heterozigotia: <b>139</b> + <b>288</b> + <b>427</b> +51+22
rs4864056		Homozigotia alelo T: <b>409</b> + <b>25</b> +214+36 Homozigotia alelo C: <b>434</b> +214+36 Heterozigotia: <b>409</b> + <b>25</b> + <b>434</b> +214+36

## TABELAS SUPLEMENTARES

Tabela suplementar 1 – Sequências dos *primers forward e reverse* utilizados nas PCRs dos SNPs a analisar por RFLP

SNP	Primers	Tm	Tamanho do fragmento amplificado
rs651539	5' ATTATTCGAGCCACCACCTG 3' (F) 5' AGCAAGGAGGGAAAGAGAGG 3' (R)	59.96°C 59.95°C	837 bp
rs2975788	5' TGTTCATGTTCCACAGGAC 3' (F) 5' GGGTGGAGACACAAAGCCTA 3' (R)	59.37°C 60.11°C	981 bp
rs17235018	5' GATGTGTCCATTGCAGGAGA 3' (F) 5' TGTCTTGGTCTCAGAAGGAA 3' (R)	59.64°C 59.83°C	500 bp
rs4864056	5' TGCTTTACCAAATGCTCTGTG 3' (F) 5' ATCAACATGCCTGCTGCTC 3' (R)	58.98°C 59.97°C	684 bp

bp – pares de bases nucleotídicas; Tm – temperatura de *melting*Tabela suplementar 2 – Sequências dos *primers forward e reverse* utilizados na PCR dos SNPs a analisar por Sequenom iPLEX

Plex	SNP	Primers	Tamanho do fragmento amplificado
1	rs17048803	5' ACGTTGGATGTCCATCAGCTGGCTATATTG 3' (F) 5' ACGTTGGATGGGAGGAGTTGTGCAAAGAAG 3' (R)	100 bp
	rs687417	5' ACGTTGGATGACAGCTGAGCAGGAATTTTG 3' (F) 5' ACGTTGGATGTCAGGAGCCCGGAGAGTCTA 3' (R)	118 bp
	rs10514631	5' ACGTTGGATGGCTCATCCTGACTTCATCAC 3' (F) 5' ACGTTGGATGTGAAGCAGGTCAGAGCTTAG 3' (R)	100 bp
	rs2699164	5' ACGTTGGATGAGTTGAGACCTATGGATAGC 3' (F) 5' ACGTTGGATGTAGCCTCAAATACATGTGGG 3' (R)	119 bp
	rs17024145	5' ACGTTGGATGTGACTGGAGCTGAGGAAAAC 3' (F) 5' ACGTTGGATGATGTCCCCATTTCCTGCAC 3' (R)	101 bp
	rs12711674	5' ACGTTGGATGCAATTGTCAAGTACATGCAC 3' (F) 5' ACGTTGGATGCTTTCATTGCGTGACAGCTC 3' (R)	107 bp
	rs1562646	5' ACGTTGGATGTCCGCTCTGTTGATGTTGG 3' (F) 5' ACGTTGGATGGCATGTTTTAATGCCTCTG 3' (R)	113 bp
	rs615329	5' ACGTTGGATGGACATGTGGGTATCTCTTGC 3' (F) 5' ACGTTGGATGAGAGTGCTCAAATTCACCTC 3' (R)	111 bp
	rs10049174	5' ACGTTGGATGTGAGCTGACTGTCTTTACAC 3' (F) 5' ACGTTGGATGCATCTCACTTGGCCTAGTC 3' (R)	111 bp
	rs11706616	5' ACGTTGGATGGGTTGCAAAGACAGGAATC 3' (F) 5' ACGTTGGATGTGTCTACACGCTTAGACCTC 3' (R)	100 bp
	rs4916566	5' ACGTTGGATGCCGTACAGGTTAACTGAAGC 3' (F) 5' ACGTTGGATGTGGGATTACAGGCGTGAGC 3' (R)	94 bp
	rs11747001	5' ACGTTGGATGCCAGGTAAATATATATCATC 3' (F) 5' ACGTTGGATGTCAAATGCAGTGGCCAGAAC 3' (R)	113 bp
	rs9350278	5' ACGTTGGATGCTTCTTGCCATGTTCTATT 3' (F) 5' ACGTTGGATGGGAAAAACATAAATGGGTG 3' (R)	119 bp
	rs9314584	5' ACGTTGGATGCCATGAAGTGGAACCATAC 3' (F) 5' ACGTTGGATGCATGCCTTCTCCACGATTAC 3' (R)	106 bp
	rs13258732	5' ACGTTGGATGAAAGGAAATAGCCTGAACAC 3' (F) 5' ACGTTGGATGCTGTTTTCACTTATCTAGCAC 3' (R)	107 bp
	rs13249775	5' ACGTTGGATGGTAAATTTATCTAAAGTAGG 3' (F) 5' ACGTTGGATGGATACTGCCTCAGAATGGTC 3' (R)	102 bp
	rs7860786	5' ACGTTGGATGTTTGTACCCCTTTCCACTC 3' (F) 5' ACGTTGGATGACTGTCCTCTCTGAAGGTC 3' (R)	95 bp

1 (cont.)	rs10903686	5' ACGTTGGATGATGACCTCATCCACAGCTAC 3' (F) 5' ACGTTGGATGGCCCCAGATTCATATGTTG 3' (R)	118 bp
	rs11001208	5' ACGTTGGATGGCACCAAACTGGTAGTCTC 3' (F) 5' ACGTTGGATGATGAGAACCTTAGTGCCTGC 3' (R)	112 bp
	rs11849729	5' ACGTTGGATGCCAGGGATCTCATAAAATCT 3' (F) 5' ACGTTGGATGGAGTGTGTTGAGATGTAACG 3' (R)	120 bp
	rs11159233	5' ACGTTGGATGAAGGACCCAGGATCTGAGAC 3' (F) 5' ACGTTGGATGTAAGCCAAAAGCCAGAGAG 3' (R)	100 bp
	rs1144046	5' ACGTTGGATGCCTTCATCCATTACACAGG 3' (F) 5' ACGTTGGATGTGGATTGGCTATTTGCAACC 3' (R)	97 bp
	rs6039208	5' ACGTTGGATGGCATGCATTAAGAGCACAG 3' (F) 5' ACGTTGGATGGGGTGTGATTACAGGATGTG 3' (R)	105 bp
	rs6086654	5' ACGTTGGATGGGCAGATCTAGACCACATTG 3' (F) 5' ACGTTGGATGACCTGACACTTGCCAATAG 3' (R)	119 bp
	rs742758	5' ACGTTGGATGCTCTTCCCCATTACACAC 3' (F) 5' ACGTTGGATGTGGCTGCTGTCTGACCTCC 3' (R)	107 bp
2	rs7514276	5' ACGTTGGATGGGGAAGAAGTCTTTCAACTG 3' (F) 5' ACGTTGGATGGGGTGTGAGATTAACAGTTAAA 3' (R)	116 bp
	rs4660095	5' ACGTTGGATGCTGACGAATCCAATACTGTTG 3' (F) 5' ACGTTGGATGGAGAAAAAGTGGCATCTTAG 3' (R)	100 bp
	rs1479249	5' ACGTTGGATGGTGACCTTAAAAAGATGAAG 3' (F) 5' ACGTTGGATGGCTCTTGAAATCTCAGGAAC 3' (R)	104 bp
	rs11933940	5' ACGTTGGATGCACCTTAGAAAGATGTACAG 3' (F) 5' ACGTTGGATGAATCACCTTGCTCTTTCC 3' (R)	88 bp
	rs2876137	5' ACGTTGGATGAACATACTTCCAAGTGATGC 3' (F) 5' ACGTTGGATGTTGTATGCTACACCCTAGAC 3' (R)	93 bp
	rs4710950	5' ACGTTGGATGAGTACTCTTCTCTTCTCTGC 3' (F) 5' ACGTTGGATGGGGCTTAGACATCTTAAATG 3' (R)	104 bp
	rs10253978	5' ACGTTGGATGTTCTGTCACAAAGCAATGCC 3' (F) 5' ACGTTGGATGAGGCATCCAAGAATTGGTAG 3' (R)	106 bp
	rs2617091	5' ACGTTGGATGTAAATACTTCTGCAGCATGG 3' (F) 5' ACGTTGGATGGCCAAAAATACATAATTAGC 3' (R)	110 bp
	rs17075289	5' ACGTTGGATGGCCCCTTAAATGAAAAGGAG 3' (F) 5' ACGTTGGATGAGATTGTGTGATAATTGAG 3' (R)	81 bp
	rs2703313	5' ACGTTGGATGGGCAGAAAGCCATGCATTTAT 3' (F) 5' ACGTTGGATGAGACTACTCAACTTTCCAGC 3' (R)	116 bp
	rs1243411	5' ACGTTGGATGGTGAATAAGGAAGCCTGGAG 3' (F) 5' ACGTTGGATGCTCCCATGACCTGATTTTTG 3' (R)	115 bp
	rs11175781	5' ACGTTGGATGTTGTGCTTGGCCTGTTTCTC 3' (F) 5' ACGTTGGATGACACTGAGCATGAAAAAAC 3' (R)	107 bp
	rs7980667	5' ACGTTGGATGTGCCCCTTTACAACCTGTCC 3' (F) 5' ACGTTGGATGCATGTCAGGATGAAGATCGG 3' (R)	106 bp
	rs16949386	5' ACGTTGGATGCTCAGAAGCCTTCTTACTGT 3' (F) 5' ACGTTGGATGGAAGTAGGTACAACTAGCAG 3' (R)	100 bp
	rs4888652	5' ACGTTGGATGAGGAGGAGATGAGCTTTTCTCAG 3' (F) 5' ACGTTGGATGGCTCAAGATCTTCCATCTTC 3' (R)	97 bp
	rs7207469	5' ACGTTGGATGCTTGAGGATTAGCTGCAGTG 3' (F) 5' ACGTTGGATGGGGTTTCATGCAGCATATAG 3' (R)	94 bp
	rs6506157	5' ACGTTGGATGTGCCATGAGCACCCTTATC 3' (F) 5' ACGTTGGATGTCCCTGTACACAAATCTTG 3' (R)	116 bp
	rs6140752	5' ACGTTGGATGTTTCTTTCATTGGCATCCTG 3' (F) 5' ACGTTGGATGCTCCTGACTTATATCTCCAC 3' (R)	97 bp
	rs5960190	5' ACGTTGGATGAGGCAAGAGTTGACGTTTGG 3' (F) 5' ACGTTGGATGTCTTTTCTGTAAACATC 3' (R)	106 bp
	rs17218233	5' ACGTTGGATGAGTTGTGCAACCAAGAACC 3' (F) 5' ACGTTGGATGGTATGAACTGAGGCTTTGTG 3' (R)	94 bp

bp – pares de bases nucleotídicas; Tm – temperatura de *melting*

**Tabela suplementar 3** – Sequências das sondas de extensão utilizadas em Sequenom iPLEX Gold

<i>Plex</i>	SNP	Sondas de Extensão
1	rs17048803	gTGTGCAAAGAAGGGATTAG
	rs687417	GGAGAGTCTACCAAGTGGA
	rs10514631	AGAGCTTAGACTTTCTCTCA
	rs2699164	tgtgtAGTTTGCCCTGGTATGT
	rs17024145	TTTCCTGCACCTTTTTC
	rs12711674	TGCGTGACAGCTCCATTTC
	rs1562646	ATTGTCTTGGCCAAAAC
	rs615329	cCTCTTCAAAGAGCCATCAAATTA
	rs10049174	ccTGCCCTAGTCATTAAACC
	rs11706616	taccCGCTTAGACCTCTCAGATA
	rs4916566	cgttAGCCACCACACTTGGCCGAT
	rs11747001	tccTCTAACAATTAACCAATCGATT
	rs9350278	ACATGTCATTCTCTACA
	rs9314584	AAAACAGACAAGAAAGCA
	rs13258732	ccccGTCTGATTACCCACTCTTAA
	rs13249775	tgaaGAAGACCTTAGTTTCAAGC
	rs7860786	tAAGGTCTAGTTGATGTAGAATT
	rs10903686	tctcCCTAATCTCCCTTGGAAC
	rs11001208	ggatAACCTTAGTGCCTGCTATTAG
	rs11849729	GAGATGTAACGTTCACTTAGC
	rs11159233	ggaCTGGTATGGCACAACCTT
	rs1144046	gggaaTGCTATTTCTATAGGCATGAT
	rs6039208	gggGATTACAGGATGTGTAATGTA
	rs6086654	tACACTTGCCAATAGTTGATATTAA
	rs742758	ctTCTGACCTCCCTTCTTAA
2	rs7514276	ccccgACACCAAATCAAACAGATTG
	rs4660095	ggtGCATCTTAGAAAGCAAATGG
	rs1479249	gggtgGAAATCTCAGGAACAGAAAAC
	rs11933940	gGCCTCTTCCCTCTAGGCTCCA
	rs2876137	ACACCCTAGACACTCCA
	rs4710950	gactaTGTGTCTGCAGAAAACTTTT
	rs10253978	CAAGAATTGGTAGTTACTATTGCC
	rs2617091	ATGCTTAATTTTTATCATGACAATA
	rs17075289	ctgGTGTGATAATTGAGATAGGTTT
	rs2703313	TCAACTTCCAGCTTGCAATTA
	rs1243411	AACAGCTCTTGCCAAAGAAGT
	rs11175781	TGAGCATGAAAAACTACAAC
	rs7980667	ggGATCGGATCAAGAAACATG
	rs16949386	AAACTGCAGTACGAGTC
	rs4888652	CTTCTTGAGATATCCTTCCC
	rs7207469	GGCAAGTTCATTAATTCAATTC

2 (cont.)	rs6506157	GGACTCATTTCAGCCAAG
	rs6140752	ACTTATATCTCCACTCTTACAC
	rs5960190	GTTTTTCTACAGAATCCTGAGTGA
	rs17218233	GCTTTGTGCTAGCAATTTA